



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 124162

TO: Rita Mitra  
Location: Rem 3 B 65  
Art Unit: 1653  
Tuesday, June 08, 2004

3C70  
236

Case Serial Number: 10/023529

From: Mary Jane Ruhl  
Location: Biotech-Chem Library  
Remsen 1-B55  
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

### Search Notes

Examiner Mitra,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl  
Technical Information Specialist  
STIC  
Remsen 1-A-62  
Ext. 22524

No art



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:35:17 ; Search time 46 Seconds  
(without alignments)  
3745.064 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 546

Sequence: 1 MKNQDKNGAKQSNPKSP.....APSTEASQTGPQEPISARA 546

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTRMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_muc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rhodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	546	100.0	546	Q86YW3	Q86YW3 homo sapien
2	529	96.9	546	Q86T54	Q86T54 homo sapien
3	433	79.3	546	Q86T86	Q86T86 homo sapien
4	243	44.5	546	Q86T85	Q86T85 homo sapien
5	116	21.2	116	Q86Y86	Q86Y86 homo sapien
6	21	3.8	150	Q8C5K1	Q8C5K1 mus musculus
7	19	3.5	241	Q8BWJ7	Q8BWJ7 mus musculus
8	19	3.5	241	Q8BPJ1	Q8BPJ1 mus musculus
9	19	3.5	463	Q8BUK2	Q8BUK2 mus musculus
10	19	3.5	524	Q8BHN1	Q8BHN1 mus musculus
11	19	3.5	528	Q8NUQ3	Q8NUQ3 homo sapien
12	19	3.5	684	Q8VBT1	Q8VBT1 mus musculus
13	17	3.1	186	Q8P0X1	Q8P0X1 homo sapien
14	17	3.1	505	Q8N3S2	Q8N3S2 homo sapien
15	17	3.1	676	Q8I969	Q8I969 gallus gall
16	17	3.1	715	Q8N3L3	Q8N3L3 homo sapien

17	17	3.1	718	4	Q86T52	Q86T52 homo sapien
18	11	2.0	404	10	Q8RXD7	Q8RXD7 arabidopsis
19	10	1.8	181	4	Q8BZA4	Q8BZA4 homo sapien
20	10	1.8	190	11	Q8BOK8	Q8BOK8 mus musculus
21	9	1.6	637	16	Q8DTCL	Q8DTCL streptococc
22	9	1.6	826	5	Q9VPS3	Q9VPS3 drosophila
23	9	1.6	842	3	Q96WP6	Q96WP6 lentinula e
24	8	1.5	113	5	Q86I22	Q86I22 dictyosteli
25	8	1.5	135	16	Q8PL61	Q8PL61 xanthomonas
26	8	1.5	135	16	Q8P9E8	Q8P9E8 xanthomonas
27	8	1.5	143	2	P70717	P70717 actinobacil
28	8	1.5	195	5	O17194	O17194 caenorhabdi
29	8	1.5	206	16	Q83C21	Q83C21 coxiella bu
30	8	1.5	224	10	Q8SCX3	Q8SCX3 arabidopsis
31	8	1.5	229	16	O60677	O60677 mycobacteri
32	8	1.5	255	16	Q8R9Z0	Q8R9Z0 thermocoaer
33	8	1.5	255	16	Q8R9J9	Q8R9J9 pseudomonas
34	8	1.5	264	5	Q86LQ1	Q86LQ1 branchiosto
35	8	1.5	281	5	O44174	O44174 caenorhabdi
36	8	1.5	312	10	Q8SZW2	Q8SZW2 arabidopsis
37	8	1.5	324	10	Q8SA04	Q8SA04 oryza sativ
38	8	1.5	346	16	Q8ELJ3	Q8ELJ3 oceanobacil
39	8	1.5	348	12	Q9Q1U7	Q9Q1U7 salmirilne
40	8	1.5	348	12	Q805G5	Q805G5 salmirilne
41	8	1.5	395	2	Q9AES1	Q9AES1 bacteroides
42	8	1.5	405	12	Q80BK0	Q80BK0 salmirilne
43	8	1.5	405	12	Q805M7	Q805M7 salmirilne
44	8	1.5	413	16	Q82DR9	Q82DR9 streptomyce
45	8	1.5	434	16	Q910M6	Q910M6 streptomyce

#### ALIGNMENTS

##### RESULT 1

Q86YW3 PRELIMINARY; PRT; 546 AA.  
 ID Q86YW3  
 AC Q86YW3; 2003 (TREMREL. 24, Created)  
 DT 01-JUN-2003 (TREMREL. 24, Last sequence update)  
 DT 01-JUN-2003 (TREMREL. 24, Last annotation update)  
 DE Taxilin.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MBLINK=22448667; PubMed=12558796;  
 RA Nogami S., Satoh S., Nakano M., Shimizu H., Fukushima H., Maruyama A.,  
 RA Terano A., Shirataki H.,  
 RT "Taxilin; a novel syntaxin-binding protein that is involved in Ca2+-  
 RT dependent exocytosis in neuroendocrine cells.";  
 RL Genes Cells 8:17-28(2003).  
 DR EMBL; AF516206; AAC42465.1;  
 SQ SEQUENCE 546 AA; 61891 MW; 698CD74F78897DF6 CRC64;

Query Match 100.0%; Score 546; DB 4; Length 546;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKNQDKNGAKQSNPKSPQPEAGPEGAQERPSQAAPAVEAEGPSSQAPRKEGAQA 60  
 DB 1 MKNQDKNGAKQSNPKSPQPEAGPEGAQERPSQAAPAVEAEGPSSQAPRKEGAQA 60  
 QY 61 RTAOSGALRDVSELSQLEDILSTYCVNNQGGEDGAGGEPAEDAEKSTTYARN 120  
 DB 61 RTAOSGALRDVSELSQLEDILSTYCVNNQGGEDGAGGEPAEDAEKSTTYARN 120  
 QY 121 GEPEPTPVNGEKEPSKGDNTETIROSDVGDHRRPQEKKAKGLGKEITLLMQTLN 180  
 DB 121 GEPEPTPVNGEKEPSKGDNTETIROSDVGDHRRPQEKKAKGLGKEITLLMQTLN 180

```

QY 161 TLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSK 240
DB 161 TLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSK 240
QY 241 LESLRELQRHNSLKEGVORAREBEERKEVTSHPQVTLNDIQLOMEQHNRNSKLQ 300
DB 241 LESLRELQRHNSLKEGVORAREBEERKEVTSHPQVTLNDIQLOMEQHNRNSKLQ 300
QY 301 ENMELAEERLKKLIEQYELREEHIDKVPFKHDLQOQLVDKLOQAQEMKAEERHOREKD 360
DB 301 ENMELAEERLKKLIEQYELREEHIDKVPFKHDLQOQLVDKLOQAQEMKAEERHOREKD 360
QY 361 FLKKEAVERSORMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWK 420
DB 361 FLKKEAVERSORMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWK 420
QY 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIORLKLALQTERND 480
DB 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIORLKLALQTERND 480
QY 481 LNKRVQDLSAGGOSLTDGSPRRPBGCAQAPSSPRVTEAPCYGAPSTASGQTGPQE 540
DB 481 LNKRVQDLSAGGOSLTDGSPRRPBGCAQAPSSPRVTEAPCYGAPSTASGQTGPQE 540
QY 541 PLSARA 546
DB 541 PLSARA 546

RESULT 2
Q86T54 PRELIMINARY; PRT; 546 AA.
AC Q86T54;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein DKFZp451K215.
GN DKFZp451K215.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Human skeletal muscle;
RA Koehler K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL812338; CAD91138.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61861 MW; 728D674F78897DF6 CRC64;

Query Match 96.9%; Score 529; DB 4; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAGKSNPKSPGQPEAGQERPSQAAPAVEAGPGSSQAQPKPEGAQA 60
DB 1 MNQDKKNGAAGKSNPKSPGQPEAGQERPSQAAPAVEAGPGSSQAQPKPEGAQA 60
QY 61 RTAQSGALRDVSELSRQLEDILSTYCVNNQGGPGEDGAQEPAPEDAEKSRITYVARN 120
DB 61 RTAQSGALRDVSELSRQLEDILSTYCVNNQGGPGEDGAQEPAPEDAEKSRITYVARN 120
QY 121 GEPEPTPVNGEKEPSKGPNTTEIRQSDVGDGRDHRPQEKKAAGLQKEITLLMQTLN 180
DB 121 GEPEPTPVNGEKEPSKGPNTTEIRQSDVGDGRDHRPQEKKAAGLQKEITLLMQTLN 180
QY 181 TLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSK 240
DB 181 TLSTPEKLAALCKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSK 240

```

```

QY 241 LESLRELQRHNSLKEGVORAREBEERKEVTSHPQVTLNDIQLOMEQHNRNSKLQ 300
DB 241 LESLRELQRHNSLKEGVORAREBEERKEVTSHPQVTLNDIQLOMEQHNRNSKLQ 300
QY 301 ENMELAEERLKKLIEQYELREEHIDKVPFKHDLQOQLVDKLOQAQEMKAEERHOREKD 360
DB 301 ENMELAEERLKKLIEQYELREEHIDKVPFKHDLQOQLVDKLOQAQEMKAEERHOREKD 360
QY 361 FLKKEAVERSORMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWK 420
DB 361 FLKKEAVERSORMCELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWK 420
QY 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIORLKLALQTERND 480
DB 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKLEGLQVKIORLKLALQTERND 480
QY 481 LNKRVQDLSAGGOSLTDGSPRRPBGCAQAPSSPRVTEAPCYGAPSTASGQTGPQS 529
DB 481 LNKRVQDLSAGGOSLTDGSPRRPBGCAQAPSSPRVTEAPCYGAPSTASGQTGPQS 529

RESULT 3
Q86T86 PRELIMINARY; PRT; 546 AA.
AC Q86T86;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein DKFZp451I0918.
GN DKFZp451I0918.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL812636; CAD89951.1; -.
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61892 MW; FEBCAD4F753F7F61 CRC64;

Query Match 79.3%; Score 433; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 13 QSNPKSPGQPEAGQERPSQAAPAVEAGPGSSQAQPKPEGAQAQARTAQSGALRDVS 72
DB 13 QSNPKSPGQPEAGQERPSQAAPAVEAGPGSSQAQPKPEGAQAQARTAQSGALRDVS 72
QY 73 BELSRQLEDILSTYCVNNQGGPGEDGAQEPAPEDAEKSRITYVARNGEPEPTPVNGE 132
DB 73 BELSRQLEDILSTYCVNNQGGPGEDGAQEPAPEDAEKSRITYVARNGEPEPTPVNGE 132
QY 133 KEPSKGPNTTEIRQSDVGDGRDHRPQEKKAAGLQKEITLLMQTLNLTSTPEKLAAL 192
DB 133 KEPSKGPNTTEIRQSDVGDGRDHRPQEKKAAGLQKEITLLMQTLNLTSTPEKLAAL 192
QY 193 CKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSKLSLCRELQHN 252
DB 193 CKKYAELEHNSQKMKLLQKQSQLVQEKDHLRGHSHKAVLARSKLSLCRELQHN 252
QY 253 RSLKEGVORAREBEERKEVTSHPQVTLNDIQLOMEQHNRNSKLQENMELAEERLKKL 312
DB 253 RSLKEGVORAREBEERKEVTSHPQVTLNDIQLOMEQHNRNSKLQENMELAEERLKKL 312
QY 313 IEQYELREEHIDKVPFKHDLQOQLVDKLOQAQEMKAEERHOREKDPLKKEAVESORM 372
DB 313 IEQYELREEHIDKVPFKHDLQOQLVDKLOQAQEMKAEERHOREKDPLKKEAVESORM 372
QY 373 CELMKQOETHLKOQALYTEKPEFQNTLSKSEVFTTFKQEMEKWKIKKLEKETTMY 432

```

Db	373	CELKMQQETHLKKQALALYTEKFEFQNTLSKSSVFVTFPQEMEKMTKKIKLEKETMY	432
Qy	433	RSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGG	492
Db	433	RSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGG	492
Qy	493	QGSILTDGSPRRPGCAQAPSSPRVTEAPCYPGAPSTASGOTGPOEPTSARA	546
Db	493	QGSILTDGSPRRPGCAQAPSSPRVTEAPCYPGAPSTASGOTGPOEPTSARA	546
RESULT 4			
ID	Q86T85	PRELIMINARY; PRT; 546 AA.	
AC	Q86T85		
DT	01-JUN-2003	(TrEMBLrel. 24, Created)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)	
DE	Hypothetical protein DKFZp451J0118.		
GN	DKFZP451J0118.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,		
RA	Osanger A., Wiemann S.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL832637; CAD89952.1; -.		
KW	Hypothetical protein.		
SQ	SEQUENCE 546 AA; 61846 MW; 36718BAE3AA7E6C2 CRC64;		
Query Match 44.5%; Score 243; DB 4; Length 546;			
Best Local Similarity 99.5%; Pred. No. 8.8e-241;			
Matches 543; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
Qy	1	MKNQDKKNGAAKQNPXSSPGQPEAGPEGAQERSQAAPAVEADPGSSQAAPKPEGAQA	60
Db	1	MKNQDKKNGAAKQNPXSSPGQPEAGPEGAQERSQAAPAVEADPGSSQAAPKPEGAQA	60
Qy	61	RTAQSGLRDVSEELSRQLEDILSTYCVDNNGGPGDGAQGRPAEPDAEKSTYVARN	120
Db	61	RTAQSGLRDVSEELSRQLEDILSTYCVDNNGGPGDGAQGRPAEPDAEKSTYVARN	120
Qy	121	GEPEPTPVNGEKEPSKGPNTTEIRQSDVGDHRRPQEKKAAGLKEITLLMOTLN	180
Db	121	GEPEPTPVNGEKEPSKGPNTTEIRQSDVGDHRRPQEKKAAGLKEITLLMOTLN	180
Qy	181	TLSTPEEKLAALCKVAFLEEHNSQKMKLLQKQSQQLVQEKDHLRGEHSAVLARSK	240
Db	181	TLSTPEEKLAALCKVAFLEEHNSQKMKLLQKQSQQLVQEKDHLRGEHSAVLARSK	240
Qy	241	LESICRELQHRNRLKEEGVQARAEERKEKKEVTSHFQVTLNDIQLQMEQHNERNSKLQ	300
Db	241	LESICRELQHRNRLKEEGVQARAEERKEKKEVTSHFQVTLNDIQLQMEQHNERNSKLQ	300
Qy	301	ENMELARLKKLIFQYELREEHIDKVKFKHDLQQLVDKLOQAQEMKAEERHOREKD	360
Db	301	ENMELARLKKLIFQYELREEHIDKVKFKHDLQQLVDKLOQAQEMKAEERHOREKD	360
Qy	361	FLKKEAVESQRMCLMKQETHLQKQALALYTEKFEFQNTLSKSSVFVTFPQEMEKMTK	420
Db	361	FLKKEAVESQRMCLMKQETHLQKQALALYTEKFEFQNTLSKSSVFVTFPQEMEKMTK	420
Qy	421	KIKLEKETMYRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERND	480
Db	421	KIKLEKETMYRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERND	480
Qy	481	LNNKRVQDLSAGGQGSILTDGSPRRRPGCAQAPSSPRVTEAPCYPGAPSTASGOTGPOE	540
Db	481	LNNKRVQDLSAGGQGSILTDGSPRRRPGCAQAPSSPRVTEAPCYPGAPSTASGOTGPOE	540

Qy	541	PTSARA 546		
Db	541	PTSARA 546		
RESULT 5				
ID	Q86Y86	PRELIMINARY; PRT; 116 AA.		
AC	Q86Y86			
DT	01-JUN-2003	(TrEMBLrel. 24, Created)		
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)		
DE	Hypothetical protein.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC046565; AAH46565.1; -.			
KW	Hypothetical protein.			
SQ	SEQUENCE 116 AA; 12623 MW; B99B79EACAAB43F CRC64;			
Query Match 21.2%; Score 116; DB 4; Length 116;				
Best Local Similarity 100.0%; Pred. No. 1.1e-110;				
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	431	MYRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSA	490	
Db	1	MYRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSA	60	
Qy	491	GGQGSILTDGSPRRRPGCAQAPSSPRVTEAPCYPGAPSTASGOTGPOEPTSARA	546	
Db	61	GGQGSILTDGSPRRRPGCAQAPSSPRVTEAPCYPGAPSTASGOTGPOEPTSARA	116	
RESULT 6				
ID	Q8C5K1	PRELIMINARY; PRT; 150 AA.		
AC	Q8C5K1			
DT	01-MAR-2003	(TrEMBLrel. 23, Created)		
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)		
DE	Hypothetical protein (fragment).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Olfactory brain;			
RX	MEDLINE=22354683; PubMed=12466851;			
RA	The PANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RT	60,770 full-length cDNAs."			
RL	Nature 420:563-573 (2002).			
DR	EMBL; AK078192; BAC37168.1; -.			
KW	Hypothetical protein.			
FT	NON TER 150 150			
SQ	SEQUENCE 150 AA; 16490 MW; C444717503E42E49 CRC64;			
Query Match 3.8%; Score 21; DB 11; Length 150;				
Best Local Similarity 100.0%; Pred. No. 8.5e-13;				
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	75	LSRQLEDILSTYCVDNNGQSP	95	
Db	66	LSRQLEDILSTYCVDNNGQSP	86	



```
Db 97 LMTLSTPEKLAALCKKYA 115

RESULT 7
Q8BWJ7 Q8BWJ7 PRELIMINARY; PRT; 241 AA.
AC Q8BWJ7
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Muscle-derived protein MDP77 variant 1 (Fragment).
GN 2310001N14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK052324; BAC34937.1; -.
DR MGD; MGI:1916756; 2310001N14RIK.
FT NON TER 241
SQ SEQUENCE 241 AA; 2732 MW; 0B2DD1BD215630BB CRC64;

Query Match 3.5%; Score 19; DB 11; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AVLARKSLKSLCRLQHN 252
|||||
DB 210 AVLARKSLKSLCRLQHN 228

RESULT 8
Q8BP11 Q8BP11 PRELIMINARY; PRT; 241 AA.
AC Q8BP11
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB3P7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mullerian duct includes surrounding region;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK078477; BAC37296.1; -.
DR MGD; MGI:194910; Rb3p7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0015664; P:transcriptional repressor activity; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P. . .; IDA.
KW Hypothetical protein.
SQ SEQUENCE 241 AA; 27598 MW; 1E0C7A38169F31A CRC64;

Query Match 3.5%; Score 19; DB 11; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 LMTLSTPEKLAALCKKYA 197
|||||

Db 97 LMTLSTPEKLAALCKKYA 115

RESULT 9
Q8BUK2 Q8BUK2 PRELIMINARY; PRT; 463 AA.
AC Q8BUK2
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Muscle-derived protein MDP77 variant 1 (Fragment).
GN 2310001N14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK084639; BAC39238.1; -.
DR MGD; MGI:1916756; 2310001N14RIK.
FT NON TER 463
SQ SEQUENCE 463 AA; 54090 MW; 18E9F881D73D58AA CRC64;

Query Match 3.5%; Score 19; DB 11; Length 463;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 234 AVLARKSLKSLCRLQHN 252
|||||
DB 210 AVLARKSLKSLCRLQHN 228

RESULT 10
Q8BHN1 Q8BHN1 PRELIMINARY; PRT; 524 AA.
AC Q8BHN1
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB3P7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex, Head, and Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK030100; BAC26785.1; -.
DR EMBL; AK031783; BAC27547.1; -.
DR EMBL; AK044130; BAC31791.1; -.
DR MGD; MGI:1134910; Rb3p7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0015664; P:transcriptional repressor activity; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P. . .; IDA.
KW Hypothetical protein.
SQ SEQUENCE 524 AA; 60308 MW; 0228777633B4ED7C CRC64;

Query Match 3.5%; Score 19; DB 11; Length 524;
Best Local Similarity 100.0%; Pred. No. 3e-10;
```

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LNTLSTPEEKLAALCKKYA 197  
Db 146 LNTLSTPEEKLAALCKKYA 164

RESULT 11

Q9NUQ3  
ID Q9NUQ3 PRELIMINARY; PRT; 528 AA.

AC	Q9NT03;
AD	
AE	
AF	
AG	
AH	
AI	
AJ	
AK	
AL	
AM	
AN	
AO	
AP	
AQ	
AR	
AS	
AT	
AV	
AW	
AX	
AY	
AZ	
BA	
BB	
BC	
BD	
BE	
BF	
BG	
BH	
BI	
BJ	
BK	
BL	
BM	
BN	
BO	
BP	
BQ	
BR	
BS	
BT	
BU	
BV	
BW	
BX	
BY	
BZ	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
C	

Query Match 3.5%; Score 19; DB 4; Length 528;  
Best Local Similarity 100.0%; Pred. No. 3e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels

Qy 179 INTLSTPEEKLAALCKYA 197  
|||  
Db 146 INTLSTPEEKLAALCKYA 164

## RESULT 12

Q8VBT1	PRELIMINARY;	684 AA.
ID	Q8VBT1	
AC	Q8VBT1;	
DT	01-MAR-2002 (TrEMBLrel. 20, Created)	
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)	
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)	
DE	Muscle-derived protein MDP77 variant 2 (Muscle-derived protein MDP77 variant 1)	
DE	231000IN14RIK OR MDP77.	
GN	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;	
OX	NCBI_TaxID=10090;	
EX	[1]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN=C57BL/6J; TISSUE=Heart;	
RC	MEDLINE=21663722; PubMed=11805083;	
RX	Benson K.F., Chada K.;	
RA	"Molecular Characterization of the Mouse In(10)17Rk Inversion and Identification of a Novel Muscle-Specific Gene at the Proximal Breakpoint.";	
RT	Genetics 160:279-287(2002).	
RL	EMBL; AF422245; AAL33910.1; -.	
RL	EMBL; AF422244; AAL33909.1; -.	
DR	MGJ; MGI:1916756; 231000IN14RIK.	
DR	SO SEQUENCE 684 AA; 77049 MW; B8A64E28EEF56831 CRC64;	

Query Match 3.5%; Score 19; DB 11; Length 684;  
Best Local Similarity 100.0%; Pred. No. 3.8e-10;  
Matches 19: Conservative 0; Mismatches 0; Indels

Qy	234	AVLARSKLES CRLQRHN	252
Db	209	AVLARSKLES CRLQRHN	227

RESULT 13

Q9POX1	PRELIMINARY;	PRT;	186 AA.
AD	Q9POX1		
AC	Q9POX1;		
DT	01-OCT-2000 (TRENBLrel. 15, Created)		
DT	01-WAR-2002 (TRENBLrel. 20, Last sequence update)		
DT	01-WAR-2002 (TRENBLrel. 20, Last annotation update)		
DE	Lipopolysaccharide specific response-5 protein (Fragment).		
GN	LSR5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RI	1;		
RP	SEQUENCE FROM N.A.		
RA	Chai Y.B., Zhao Z.L., Zhu P., Yan W., Chen N.C., Wang Q., Yue L.,		
RA	Chen S.M.;		
RT	"New Homo sapiens gene from dental pulp cells.";		
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; AF13740; AAF70546.2; -.		
RL	1		
FT	NON TER		
FT	1		
SO	SEQUENCE	186 AA; 21464 MW; 93BF83838096B0A	CRC64;

Query Match 3.1%; Score 17; DB 4; Length 186;  
Best Local Similarity 100.0%; Pred. No. 1.4e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels

Qy 413 QEMEKTKKIKLEKET 429  
|||  
nb 38 QEMEKTKKIKLEKET 54

## RESULT 14

Q8N3S2	PRELIMINARY;	PRT;	505 AA.
ID	Q8N3S2		
AC	Q8N3S2;		
DT	01-OCT-2002 (TREMBLrel. 22, Created)		
DT	01-OCT-2002 (TREMBLrel. 22; last sequence update)		
DT	01-OCT-2002 (TREMBLrel. 22; last annotation update)		
DE	Hypothetical protein (Fragment).		
DE	Hypothetical protein (Fragment).		
GN	DKFZP451A175.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
XP	[1]		
RP	SEQUENCE FROM N.A.		
RA	Koehrer K., Beyer A., Mewes H.W., Weil B., Wiemann S.;		
RA	Submitted (JUL1 2002) to the EMBL/GenBank/DBJ Databases.		
RL	EMBL; AL832322; CAD38617.1; -.		
DR	Hypothetical protein.		
KW	Hypothetical protein.		
FT	NOT RECORDED		
ST	SEQUENCE		
SV	505 AA; 56957 MW; 1AA33548A2F1DADE CRC64;		

Query Match	3.1%;	Score 17;	DB 4;	Length 505;
Best Local Similarity	100.0%;	Pred. No. 3.3e-08;		
Matches 17	Conservation 0;	Mismatches 0;	Indels 0;	

QY 236 LARSKLESICRELQRHN 252  
 |||||  
 30 LARSKLESICRELQRHN 46

## RESULT 15

Q9I969  
ID Q9I969 PRELIMINARY; PRT; 676 AA.  
AC Q9I969;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Muscle derived protein.  
GN MDP77.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
R2 SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RX MEDLINE=20175243; PubMed=10708594;  
RA Uyeda A.; Fukui I.; Fujimori K.; Kiyosue K.; Nishimune H.; Kasai M.,  
RA Taguchi T.;  
RT "MDP77: A novel neurite-outgrowth-promoting protein predominantly  
RT expressed in chick muscles.";  
RL Biochem. Biophys. Res. Commun. 269:564-569(2000).  
DR EMBL; D89999; BAA94755.1; -.  
DR FIR; JC7222; JC7222.  
SQ SEQUENCE 676 AA; 77020 MW; FCEA9E393250EE94 CRC64;  
  
Query Match 3.1%; Score 17; DB 13; Length 676;  
Best Local Similarity 100.0%; Pred. No. 4.3e-08;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 236 LARSKLESICRELOHNN 252  
DB 207 LARSKLESICRELOHNN 223  
|||||  
|||||

Search completed: June 8, 2004, 16:39:31  
Job time : 58 secs

	Query Match	100.0%;	Score 546;	DB 4;	Length 546;	
	Best Local Similarity	100.0%;	Pred. No. 0;			
	Matches 546; Conservative	0;	Mismatches	0;	Indels	Gaps 0;
OY	1	MKNQDKNGAAKOSNPKSSPGQPEAGPFGAQRPSQAAPVAEABGPGSQARPKPEGAAQ	60			
Ddb	1	MKNQDKNGAAKAYSNPKSSPGQPEAGPFGAQRPSQAAPVAEABGPGSQARPKPEGAAQ	60			
OY	61	RTAQSGALRDVSBEHSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKSRTTVARN	120			
Ddb	61	RTAQSGALRDVSBEHSRQLEDILSTYCVNNQGGPGEDGAQGEPAEPEDAERKSRTTVARN	120			
OY	121	GEPEPTPVNCEKBSPKSDPMTEELRQDSVDGDRHRRPOEKKAKGLCKETILLMOTLN	180			
Ddb	121	GEPEPTPVNCEKEBSPKSDPMTEELRQDSVDGDRHRRPOEKKAKGLCKETILLMOTLN	180			
OY	181	TLSTPEEKLAALCKKYAEELLBEHRNSQOMKLLOKQSOLVQEKDHLRGESKAVLARSK	240			
Ddb	181	TLSTPEEKLAALCKKYAEELLBEHRNSQOMKLLOKQSOLVQEKDHLRGESKAVLARSK	240			

```

Query: Maccn
      100.0%; Score: 100.0%; E: 0.000000; Length: 10
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1	MKNQDKNGAAKQSNPKSSPGQPPAGPEGAGQERPSQAAPAVEABGPGSSQAQPKRPEGAQA	60
	db	
1	MKNQDKNGAAKQSNPKSSPGQPPAGPEGAGQERPSQAAPAVEABGPGSSQAQPKRPEGAQA	60
	db	
61	RTAQSGALRVSELSQLIEDILSTYCVDNNGGPGEDGAQGBPAEPDEAKSSTYYARN	120
	db	
61	RTAQSGALRVSELSQLIEDILSTYCVDNNGGPGEDGAQGBPAEPDEAKSSTYYARN	120
	db	
121	GEPSPTPVNGEKEPSKGDPNTEIROSDEVDGRDHRPQEKKAAGLKGKEITLLMQTLN	180
	db	
121	GEPSPTPVNGEKEPSKGDPNTEIROSDEVDGRDHRPQEKKAAGLKGKEITLLMQTLN	180
	db	
181	TLSTPEKLAALCKKYAELEFEHNSSOKMKLQKQSQSLVQEKDHLRGHSKAVLARSK	240
	db	
181	TLSTPEKLAALCKKYAELEFEHNSSOKMKLQKQSQSLVQEKDHLRGHSKAVLARSK	240
	db	

Result No.	Score	Query Match	Length	DB	ID	Description
1	546	100.0	546	4	US-09-616-289-44	Sequence 44, Appl
2	429	78.6	530	4	US-08-978-608A-8	Sequence 8, Appl
3	429	78.6	530	4	US-09-517-849-8	Sequence 8, Appl
4	423	78.6	530	4	US-09-616-289-8	Sequence 8, Appl
5	172	31.5	557	4	US-08-978-608A-5	Sequence 5, Appl
6	172	31.5	557	4	US-09-517-849-5	Sequence 5, Appl
7	172	31.5	557	4	US-09-616-289-5	Sequence 5, Appl
8	19	3.5	386	2	US-08-968-751-6	Sequence 6, Appl
9	8	1.5	165	4	US-09-253-931A-19490	Sequence 19490, A
10	8	1.5	260	2	US-08-879-561-7	Sequence 7, Appl
11	7	1.3	7	4	US-08-978-608A-41	Sequence 41, Appl
12	7	1.3	7	4	US-09-517-849-41	Sequence 41, Appl
13	7	1.3	7	4	US-09-616-289-41	Sequence 41, Appl
14	7	1.3	64	4	US-09-489-039A-9565	Sequence 9565, Ap
15	7	1.3	73	3	US-09-146-950-25	Sequence 25, Appl
16	7	1.3	105	4	US-09-732-210-434	Sequence 434, App
17	7	1.3	124	4	US-09-489-039A-10493	Sequence 10493, A
18	7	1.3	127	4	US-08-849-303-19	Sequence 19, Appl
19	7	1.3	142	4	US-09-489-039A-9418	Sequence 9418, Ap
20	7	1.3	159	3	US-09-146-950-20	Sequence 20, Appl
21	7	1.3	177	4	US-09-079-030-117	Sequence 117, App
22	7	1.3	182	4	US-09-328-352-7281	Sequence 7281, Ap
23	7	1.3	197	3	US-09-146-950-18	Sequence 18, Appl
24	7	1.3	211	4	US-09-134-001C-5175	Sequence 5175, Ap
25	7	1.3	261	2	US-08-879-561-1	Sequence 1, Appl
26	7	1.3	267	4	US-09-489-039A-13457	Sequence 13457, A
27	7	1.3	274	1	US-08-287-959-3	Sequence 3, Appl



Query Match 78.6%; Score 429; DB 4; Length 530;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 17 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSSQAAPRKEGAQARTAGSALRDVSEELS 76  
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSSQAAPRKEGAQARTAGSALRDVSEELS 60  
QY 77 RQEDILSTYCVNNQGGPGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136  
DB 61 RQEDILSTYCVNNQGGPGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120  
QY 137 KQDPNTEIRQSDVEGDRHRRPOEKKKAGLGEITLLMOTLNTLSTPEKLAALCKKY 196  
DB 121 KQDPNTEIRQSDVEGDRHRRPOEKKKAGLGEITLLMOTLNTLSTPEKLAALCKKY 180  
QY 197 AELLBEHRNSQKMKLQKQSQVLQVQKDLRGEHSKAVLARSKLESCLRELQHRNRSK 256  
DB 181 AELLBEHRNSQKMKLQKQSQVLQVQKDLRGEHSKAVLARSKLESCLRELQHRNRSK 240  
QY 257 EGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 316  
DB 241 EGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300  
QY 317 ELREEHIDKVFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376  
DB 301 ELREEHIDKVFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360  
QY 377 KQETHLKQQLALYTEKFEFQNTLSKSEVFTTFKQEMEKWTCKIKLEKETTWYRSRW 436  
DB 361 KQETHLKQQLALYTEKFEFQNTLSKSEVFTTFKQEMEKWTCKIKLEKETTWYRSRW 420  
QY 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSGL 496  
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSGL 480  
QY 497 TDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 546  
DB 481 TDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530

RESULT 5  
US-08-979-608A-5  
; Sequence 5, Application US/08979608A  
; Patent No. 6355451  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; Lees, Robert S.  
; Law, Simon W.  
; Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
; TREATING ATHEROSCLEROSIS  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fish & Richardson P.C.  
; STREET: 225 Franklin Street

FILING DATE: 26-NOV-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: 10797-003001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 530 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-517-849-8

Query Match 78.6%; Score 429; DB 4; Length 530;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 17 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSSQAAPRKEGAQARTAGSALRDVSEELS 76  
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEABGPGSSQAAPRKEGAQARTAGSALRDVSEELS 60  
QY 77 RQEDILSTYCVNNQGGPGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136  
DB 61 RQEDILSTYCVNNQGGPGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120  
QY 137 KQDPNTEIRQSDVEGDRHRRPOEKKKAGLGEITLLMOTLNTLSTPEKLAALCKKY 196  
DB 121 KQDPNTEIRQSDVEGDRHRRPOEKKKAGLGEITLLMOTLNTLSTPEKLAALCKKY 180  
QY 197 AELLBEHRNSQKMKLQKQSQVLQVQKDLRGEHSKAVLARSKLESCLRELQHRNRSK 256  
DB 181 AELLBEHRNSQKMKLQKQSQVLQVQKDLRGEHSKAVLARSKLESCLRELQHRNRSK 240  
QY 257 EGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 316  
DB 241 EGVQARBEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENNELAERLKKLIEQY 300  
QY 317 ELREEHIDKVFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376  
DB 301 ELREEHIDKVFKHDLQQLVDKQLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360  
QY 377 KQETHLKQQLALYTEKFEFQNTLSKSEVFTTFKQEMEKWTCKIKLEKETTWYRSRW 436  
DB 361 KQETHLKQQLALYTEKFEFQNTLSKSEVFTTFKQEMEKWTCKIKLEKETTWYRSRW 420  
QY 437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSGL 496  
DB 421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSGL 480  
QY 497 TDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 546  
DB 481 TDSGPERRPBGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQPTTSARA 530

RESULT 4  
US-09-517-849-8  
; Sequence 8, Application US/09616289  
; Patent No. 6632523  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/09/616,289  
; CURRENT FILING DATE: 2000-07-14

```
;
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,608A
; FILING DATE: 26-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,547
; FILING DATE: 03-JUN-1997
; APPLICATION NUMBER: US 60/031,930
; FILING DATE: 27-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-979-608A-5

Query Match 31.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 5.2e-150;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 TLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQVELREEHIDKVFKHKLQQLQVLDA 339
DB 280 TLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQVELREEHIDKVFKHKLQQLQVLDA 339
QY 340 KLOQAQEMLKEAEERHQRKDFLLKEAVESQRMCELMKQETHLKKQALALYTKPEFQON 399
DB 340 KLOQAQEMLKEAEERHQRKDFLLKEAVESQRMCELMKQETHLKKQALALYTKPEFQON 399
QY 400 TLSKSESVFTTFKQEMERKTKIKLEKETTMYRSRWESSNKALLEABEKT 451
DB 400 TLSKSESVFTTFKQEMERKTKIKLEKETTMYRSRWESSNKALLEABEKT 451

RESULT 6
US-09-517-849-5
; Sequence 5, Application US/09517849
; Patent No. 6605588
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
; TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/517,849
; FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/979,608
; FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 557 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-517-849-5

Query Match 31.5%; Score 172; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 5.2e-150;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 TLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQVELREEHIDKVFKHKLQQLQVLDA 339
DB 280 TLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQVELREEHIDKVFKHKLQQLQVLDA 339
QY 340 KLOQAQEMLKEAEERHQRKDFLLKEAVESQRMCELMKQETHLKKQALALYTKPEFQON 399
DB 340 KLOQAQEMLKEAEERHQRKDFLLKEAVESQRMCELMKQETHLKKQALALYTKPEFQON 399
QY 400 TLSKSESVFTTFKQEMERKTKIKLEKETTMYRSRWESSNKALLEABEKT 451
DB 400 TLSKSESVFTTFKQEMERKTKIKLEKETTMYRSRWESSNKALLEABEKT 451

RESULT 7
US-09-616-289-5
; Sequence 5, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-09-616-289-5
```

Query Match 31.5%; Score 172; DB 4; Length 557;  
Best Local Similarity 100.0%; Pred. No. 5.2e-150;  
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 280 TLNDIQLQMHQHNRSKLRQENMELARLKLLIEQVELREHIDKVPKHDIQQQLVDA 339  
Db 280 TLNDIQLQMHQHNRSKLRQENMELARLKLLIEQVELREHIDKVPKHDIQQQLVDA 339

Qy 340 KLAQAQMLKAEERHQRKDFLLKEAVESQRCMLKQOETHLKOQLALYTKPEFQ 399  
Db 340 KLAQAQMLKAEERHQRKDFLLKEAVESQRCMLKQOETHLKOQLALYTKPEFQ 399

Qy 400 TLKSSSVFTFKQEKEMTKKKLEKETMYRSRWSSNKALLEAREKT 451  
Db 400 TLKSSSVFTFKQEKEMTKKKLEKETMYRSRWSSNKALLEAREKT 451

RESULT 8  
US-08-968-751-6  
; Sequence 6, Application US/08968751  
; Patent No. 5948643  
; GENERAL INFORMATION:  
; APPLICANT: Rubinfeld, Bonnie  
; APPLICANT: Polakis, Paul G.  
; APPLICANT: Ligenfelter, Carol  
; APPLICANT: Vuong, Tetyln T.  
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ONYX Pharmaceuticals, Inc.  
; STREET: 3031 Research Drive  
; CITY: Richmond  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94806  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968,751  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Giotta, Gregory  
; REGISTRATION NUMBER: 32,028  
; REFERENCE/DOCKET NUMBER: ONYX1024 GG  
; TELEPHONE: (510) 262-8710  
; TELEFAX: (510) 222-9758  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 386 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-968-751-6

Query Match 3.5%; Score 19; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 2.7e-09;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LNTLSTPEEKLAALCKKYA 197  
Db 135 LNTLSTPEEKLAALCKKYA 153

RESULT 9  
US-09-252-991A-19490  
; Sequence 19490, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19490  
; LENGTH: 165  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; US-09-252-991A-19490

Query Match 1.5%; Score 8; DB 4; Length 165;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 153 DRDHRPQ 160  
Db 40 DRDHRPQ 47

RESULT 10  
US-08-879-561-7  
; Sequence 7, Application US/08879561  
; Patent No. 5817482  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/879,561  
; FILING DATE: Herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0325 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 260 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:



```
; LIBRARY: GenBank
; CLONE: 1480198
; US-08-879-561-7

Query Match      1.5%; Score 8; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 261 QRAEEEE 268
Db 188 QRAEEEE 195
|||||

RESULT 11
US-08-979-608A-41
; Sequence 41, Application US/08979608A
; Patent No. 6355451
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;               Lees, Robert S.
;               Law, Simon W.
;               Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN
;                   BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND
;                   TREATING ATHEROSCLEROSIS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/517,849
;   FILING DATE: 02-Mar-2000
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/979,608
;   FILING DATE: 26-NOV-1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Myers, Louis
;   REGISTRATION NUMBER: 35,965
;   REFERENCE/DOCKET NUMBER: 10797-003001
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/542-5070
;   TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 41:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-517-849-41

Query Match      1.3%; Score 7; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 RDVSEEL 75
Db 1 RDVSEEL 7
|||||

RESULT 13
US-09-616-289-41
; Sequence 41, Application US/09616289
; Patent No. 6632923
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
;               Lees, Robert S.
;               Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
;                   PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
;                   ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
```

QY 161 EKKKAG 167  
 Db 71 EKKKAG 77  
 Search completed: June 8, 2004, 16:40:36  
 Job time : 30 secs

PRIOR APPLICATION NUMBER: US 60/031,930  
 PRIOR FILING DATE: 1996-11-27  
 PRIOR APPLICATION NUMBER: US 60/048,547  
 PRIOR FILING DATE: 1997-06-03  
 NUMBER OF SEQ ID NOS: 53  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 41  
 LENGTH: 7  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-616-289-41

Query Match 1.3%; Score 7; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 3e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 RDVSEEL 75  
 Db 1 RDVSEEL 7

RESULT 14  
 US-09-489-039A-9565  
 Sequence 9565, Application US/09489039A  
 Patent No. 6610836  
 GENERAL INFORMATION:  
 APPLICANT: Gary Breton et. al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 2709.2004001  
 CURRENT APPLICATION NUMBER: US/09/489,039A  
 CURRENT FILING DATE: 2000-01-27  
 PRIOR APPLICATION NUMBER: US 60/117,747  
 PRIOR FILING DATE: 1999-01-29  
 NUMBER OF SEQ ID NOS: 14342  
 SEQ ID NO 9565  
 LENGTH: 64  
 TYPE: PRT  
 ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-9565

Query Match 1.3%; Score 7; DB 4; Length 64;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 VVNGEKE 134  
 Db 39 VVNGEKE 45

RESULT 15  
 US-09-146-950-25  
 Sequence 25, Application US/09146950A  
 Patent No. 6287808  
 GENERAL INFORMATION:  
 APPLICANT: Busfield, Samantha J.  
 TITLE OF INVENTION: NOVEL MOLECULES OF THE HERPESVIRUS-ENTRY-MEDIATOR-RELATED  
 TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF  
 FILE REFERENCE: 09404/057601  
 CURRENT APPLICATION NUMBER: US/09/146,950A  
 CURRENT FILING DATE: 1998-09-03  
 NUMBER OF SEQ ID NOS: 25  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 25  
 LENGTH: 77  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-146-950-25

Query Match 1.3%; Score 7; DB 3; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:33:31 ; Search time 60 Seconds  
(without alignments)  
2571.182 Million cell updates/sec

Title: US-10-023-529-44  
Perfect score: 546  
Sequence: 1 MKNQDKKNGAAGQSNPKSP.....APSTASGGTGQPEPTSARA 546

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	546	100.0	546	4 AAB82808	Human low
2	546	100.0	639	6 ABR41210	Human DIT
3	546	100.0	639	6 ABR41186	Human DIT
4	476	87.2	510	4 ABB11764	Human LDL
5	476	87.2	510	4 AAM79741	Human pro
6	429	78.6	530	2 AAW49042	Human low
7	429	78.6	530	4 AAB82804	Human low
8	356	65.2	356	6 ABO07223	Human p53
9	347	63.6	358	4 AAM78757	Human pro
10	228	41.8	1749	4 AEG00839	Novel hum
11	172	31.5	557	2 AAW49039	Rabbit lo
12	172	31.5	557	4 AAB82801	Rabbit lo
13	172	31.5	557	4 AAB82801	Human bre
14	19	3.5	385	2 AAW37883	BRCA1 mod
15	19	3.5	386	2 AAY30151	Amino aci
16	19	3.5	395	6 ABO07222	Human p53
17	19	3.5	528	4 AAB31594	Human pro
18	17	3.1	85	5 ABR42281	Human ova
19	17	3.1	676	2 AAY17863	Neurite e
20	17	3.1	684	6 ABR43284	Human neu
21	12	2.2	475	4 AAO08301	Human pol
22	12	2.2	475	7 ADC33216	Human nov
23	9	1.6	99	4 AAW23308	Human car
24	9	1.6	99	7 ADE46276	Human car
25	9	1.6	554	4 ABB59454	Drosophila

26	9	1.6	637	5 ABB44191	Protein e
27	9	1.6	729	2 AAR89275	Yeast coa
28	8	1.5	65	2 AAY12665	Human 5'
29	8	1.5	156	2 AAY01195	Polypepti
30	8	1.5	170	3 AAG08727	Arabidops
31	8	1.5	174	3 AAG08726	Arabidops
32	8	1.5	194	4 AAG19843	Novel hum
33	8	1.5	224	3 AAG20570	Arabidops
34	8	1.5	229	5 ABO05655	M. tuberc
35	8	1.5	360	5 ABO07743	Selected
36	8	1.5	360	5 ABO07743	Selected
37	8	1.5	364	5 ABB11114	Yeast sel
38	8	1.5	388	7 ADB95026	A. thalia
39	8	1.5	416	7 ADC31613	Human nov
40	8	1.5	484	4 AAB79356	Corynebac
41	8	1.5	484	4 AAG92876	C. glutami
42	8	1.5	785	3 AAG41279	Arabidops
43	8	1.5	823	3 AAG41278	Arabidops
44	8	1.5	836	3 AAG41277	Arabidops
45	8	1.5	952	7 ADB95084	A. thalia

ALIGNMENTS

RESULT 1  
AAB82808  
ID AAB82808 standard; protein; 546 AA.  
XX  
AC AAB82808;  
DT 12-NOV-2001 (first entry)  
XX  
DE Human low density lipoprotein binding protein 3 (LBP-3).  
XX  
XX Low density lipoprotein binding protein 3; LBP-3; LDL; human;  
KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200164874-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 28-FEB-2001; 2001WO-US006355.  
XX  
PR 02-MAR-2000; 2000US-00517849.  
PR 14-JUL-2000; 2000US-00616289.  
XX  
(BOST-) BOSTON HEART FOUND INC.  
PI Lees AM, Lees RS, Law SW, Arjona AA;  
XX  
WPI; 2001-565505/63.  
XX  
N-PSDB; AAB26501.  
XX  
New isolated low density lipoprotein binding polypeptide for treating,  
diagnosing and/or identifying therapeutic agents for atherosclerosis.  
XX  
Claim 13(1); Fig 8A; 143pp; English.

The present sequence is that of novel human low density lipoprotein binding protein 3 (LBP-3). The amino acid sequence was deduced from the coding region of isolated genomic DNA (see AAB26501). It differs from the sequence predicted from an LBP-3 cDNA clone (see AAB2804) by the presence of an additional 16 amino acids at the N-terminus (the cDNA clone is 5' truncated) and by having asparagine at amino acid position 130 rather than tyrosine. Human LBP-2 is an example of claimed LBP polypeptides of the invention that are capable of binding to native and methylated low density lipoproteins. Also claimed are biologically active fragments and analogues of these LBPs, polynucleotides encoding LBPs, as well as expression vectors, cells and methods of producing the LBPs. A polypeptide having the amino acid residues 96-110 of the present sequence



PR	20-JUN-2001; 2001US-0299776P.
PR	20-JUN-2001; 2001US-0300001P.
XX	(INCY-) INCYTE GENOMICS INC.
PI	Daffo A, Jones AL, Tran AB, Bahl CR, Gietzen D, Chinn J;
PI	Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;
PI	Daughterly SC, Dam TC, Liu TP, Nguyen DA, Kleefeld Y, Gerstin EH;
PI	Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI	Flores V, Marwaha R, Lo A, Lan RY, Urashka MB;
XX	WPI; 2003-129518/12.
DR	N-PSDB; ACC46130.
XX	Novel human diagnostic and therapeutic polypeptide useful for identifying
PT	test compound which specifically binds to a polypeptide encoded by human
PT	diagnostic and therapeutic polynucleotide, and to induce antibodies.
PS	Claim 27; SEQ ID NO 721; 591pp; English.
XX	The invention relates to novel human diagnostic and therapeutic
CC	polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC	proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC	polynucleotide sequences at least 90% identical to the dithp cDNA
CC	sequences of the invention; recombinant vectors, host cells and
CC	transgenic organisms comprising a dithp nucleic acid sequence; the
CC	recombinant production of DITHP proteins; antibodies specific for DITHP
CC	proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC	detecting dithp nucleotide and protein sequences; methods of screening
CC	for compounds which specifically bind a DITHP protein; and methods of
CC	assessing the toxicity of test compounds using a dithp hybridisation
CC	probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC	diagnosis of a wide variety of conditions including cancer and other cell
CC	proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC	viral, fungal or parasitic infections; hormonal disorders; metabolic
CC	disorders; neurological disorders; gastrointestinal disorders; transport
CC	disorders; and connective tissue disorders. They may also be used to
CC	screen for modulators of protein activity or gene expression. DITHP
CC	proteins can additionally be used in analysis of the proteome of a tissue
CC	or cell type and to induce antibodies. The dithp nucleic acids are
CC	additionally useful in somatic or germline gene therapy of the disorders
CC	mentioned above, as a source of antisense sequences, as a source of
CC	probes and primers, in genotyping and identification of individuals, in
CC	the generation of transgenic animal models of human disease or knock in
CC	humanised animals, in toxicological testing, and in transcript imaging.
CC	The present sequence represents a DITHP protein which has extracellular
CC	signalling activity. Note: The sequence data for this patent did not form
CC	part of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 639 AA;
SQ	Query Match 100.0%; Score 546; DB 6; Length 639;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MKNQDKNGAAKQSNPKSPGQPEAGPEGQERPSQAAPAVEAEGPGSSQAAPKPEGAQA 60
DB	94 MKNQDKNGAAKQSNPKSPGQPEAGPEGQERPSQAAPAVEAEGPGSSQAAPKPEGAQA 153
QY	61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGQGPEDGAGQEPAPDAEKSRITVARN 120
DB	154 RTAQSGALRDVSEELSRQLEDILSTYCVDNNGQGPEDGAGQEPAPDAEKSRITVARN 213
QY	121 GEPEPTPVNGEKEPSKGDNTBEIROSDEVGDGRDHRPPEKKAKGLGKEITLLMOTLN 180
DB	214 GEPEPTPVNGEKEPSKGDNTBEIROSDEVGDGRDHRPPEKKAKGLGKEITLLMOTLN 273
QY	181 TLSTPEEKLAALCKKVAELLEHRNSQKMKLQKQSOVLVQEKDHLRGEHSAVLARSK 240
DB	274 TLSTPEEKLAALCKKVAELLEHRNSQKMKLQKQSOVLVQEKDHLRGEHSAVLARSK 333
QY	241 LESICRELQHRNLSKEEGVQARSEEEKRKEVTSHQVTLNDIQLQMEQHNRNSKLQR 300

Db 334 LESLRLQNRRLKEGVQARBEKEKREVTSHFQVTLNDIQLQMEHNRNSKLRQ 393  
Qy 301 ENMELARLKKLIHQYELRBEHIDKVFHKDLQQLVDKLAQQAQEKLEAEERHOREKD 360  
Db 394 ENMELARLKKLIHQYELRBEHIDKVFHKDLQQLVDKLAQQAQEKLEAEERHOREKD 453  
Qy 361 FLKKEAVESQRMCELMKQOETHLKOQALYTEKPEEPONTLSKSEVFTTFKEMEXOTK 420  
Db 454 FLKKEAVESQRMCELMKQOETHLKOQALYTEKPEEPONTLSKSEVFTTFKEMEXOTK 513  
Qy 421 KIKKLEKETTMYRSMSSNKALLEAEKTVRDKEGLQVKIQRLKLCRALQYERND 490  
Db 514 KIKKLEKETTMYRSMSSNKALLEAEKTVRDKEGLQVKIQRLKLCRALQYERND 573  
Qy 481 LNKGVQDLSAGQSLTDSGPERPPEGPAQSPSPRVTEAPCYPGAPSTPAGSQTPQE 540  
Db 574 LNKGVQDLSAGQSLTDSGPERPPEGPAQSPSPRVTEAPCYPGAPSTPAGSQTPQE 633  
Qy 541 PTSARA 546  
Db 634 PTSARA 639  
RESULT 4  
ID ABB11764  
XX ABB11764 standard; peptide; 510 AA.  
AC ABB11764;  
XX 11-JAN-2002 (first entry)  
DT Human LDL binding protein homologue, SEQ ID NO:2134.  
DE Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; anti-inflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulnery; antiulcer.  
OS Homo sapiens.  
XX WO200157188-A2.  
PN 09-AUG-2001.  
XX 05-FEB-2001; 2001WO-US003800.  
XX 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
XX (HYSE-) HYSEQ INC.  
PA Tang YT, Liu C, Drmanac RT;  
PI WPI; 2001-457740/49.  
DR N-PSDB; ABA09008.  
XX Human proteins and DNA encoding sequences useful for preventing, treating  
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis  
PT and cancer.  
XX Claim 20; Page 240-241; 1963pp; English.  
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and

CC sequences ABA09225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis  
CC depending on their biological activities. Polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
CC polypeptide of the invention  
XX  
SQ Sequence 510 AA;  
Query Match 87.2%; Score 476; DB 4; Length 510;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MNQDKKNGAARQSNPKSSPGQPEAGPAQRRPSQAPAVAEFGPGSSQAPRKPGAQA 60  
Db 12 MNQDKKNGAARQSNPKSSPGQPEAGPAQRRPSQAPAVAEFGPGSSQAPRKPGAQA 71  
Qy 61 RTAQSGALRDVSELSRQLEDILSTYCVDNNGGPGEDGAQGEPAEPDAEKSRITYARN 120  
Db 72 RTAQSGALRDVSELSRQLEDILSTYCVDNNGGPGEDGAQGEPAEPDAEKSRITYARN 131  
Qy 121 GEPEPTFVNGKEKSGDPNTEETROSDEVDGRDHRPQEKKKAGLCKEITLLMOTLN 180  
Db 132 GEPEPTFVNGKEKSGDPNTEETROSDEVDGRDHRPQEKKKAGLCKEITLLMOTLN 191  
Qy 181 TLSTPEEKLAALCKKYAEELLEHRNSQKMLLQKQSOVLQVKDHLRGEHSKAVLARSK 240  
Db 192 TLSTPEEKLAALCKKYAEELLEHRNSQKMLLQKQSOVLQVKDHLRGEHSKAVLARSK 251  
Qy 241 LESLCHLQHRNLSLKEGVQARBEKEKREVTSHFQVTLNDIQLQMEHNRNSKLRQ 300  
Db 252 LESLCHLQHRNLSLKEGVQARBEKEKREVTSHFQVTLNDIQLQMEHNRNSKLRQ 311  
Qy 301 ENMELARLKKLIHQYELRBEHIDKVFHKDLQQLVDKLAQQAQEKLEAEERHOREKD 360  
Db 312 ENMELARLKKLIHQYELRBEHIDKVFHKDLQQLVDKLAQQAQEKLEAEERHOREKD 371  
Qy 361 FLKKEAVESQRMCELMKQOETHLKOQALYTEKPEEPONTLSKSEVFTTFKEMEXOTK 420  
Db 372 FLKKEAVESQRMCELMKQOETHLKOQALYTEKPEEPONTLSKSEVFTTFKEMEXOTK 431  
Qy 421 KIKKLEKETTMYRSMSSNKALLEAEKTVRDKEGLQVKIQRLKLCRALQYERND 476

Db 432 KIKLEKETTMYRSWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQT 487

RESULT 5

AAW79741

ID AAW79741 standard; protein; 510 AA.

XX AC AAW79741;

DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3387.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX WO200157190-A2.

PN 09-AUG-2001.

PD 05-FEB-2001; 2001WO-US004098.

PF 03-FEB-2000; 2000US-00495914.

XX PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR N-PSDB; AAK52874.

DR Nucleic acids encoding polypeptides with cytokine-like activities, useful

PT in diagnosis and gene therapy.

XX Claim 20; Page 319; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111

CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the

XX sequence listing were missing at the time of publication

XX Sequence 510 AA;

Query Match 87.2%; Score 476; DB 4; Length 510;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNQDKNGAAGKSNPKSSQPEAGPEGAQEPQAPQAPAVEAGPGSSQAPKPEGAQA 60

Db 12 MKNQDKNGAAGKSNPKSSQPEAGPEGAQEPQAPQAPAVEAGPGSSQAPKPEGAQA 71

QY 61 RTAQSGALRDVSEBELSROLEDILSTYCVNNDNQGPGEDGAQGEPAEPDAEKSRITYARN 120

Db 72 RTAQSGALRDVSEBELSROLEDILSTYCVNNDNQGPGEDGAQGEPAEPDAEKSRITYARN 131

QY 121 GEPEPTPVNGEKEPSKGDPTNTEBIROSDEVDGDRHRRPOEKKAKAGLGKGTITLLMOTLN 180

Db 132 GEPEPTPVNGEKEPSKGDPTNTEBIROSDEVDGDRHRRPOEKKAKAGLGKGTITLLMOTLN 191

QY 181 TLSTPEEKLAALCKKYABLLBEHRNSQOMKLLQKOSQOLVOEKDHLRGEHSKAVLAASK 240

Db 192 TLSTPEEKLAALCKKYABLLBEHRNSQOMKLLQKOSQOLVOEKDHLRGEHSKAVLAASK 251

QY 241 LSSICRELQRNRSLSKEGVORAREEERKREKRVTSHTFOVTLNDICLOWEQHNRNSKLRO 300

Db 252 LESICRELQRNRSLSKEGVORAREEERKREKRVTSHTFOVTLNDICLOWEQHNRNSKLRO 311

QY 301 ENMELAEKLLKLIQYELREBEHIDKVFHKDLQOOLVDAKLQOAEMLKZABERHOREKD 360

Db 312 ENMELAEKLLKLIQYELREBEHIDKVFHKDLQOOLVDAKLQOAEMLKZABERHOREKD 371

QY 361 FILKEAVESQRCMLMKQOETHLKQOLALYTEKEEFPONTLSKSEVPTTKQEMEKMTK 420

Db 372 FILKEAVESQRCMLMKQOETHLKQOLALYTEKEEFPONTLSKSEVPTTKQEMEKMTK 431

QY 421 KIKLEKETTMYRSWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQT 476

Db 432 KIKLEKETTMYRSWESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQT 487

RESULT 6

AAW49042

ID AAW49042 standard; protein; 530 AA.

XX AC AAW49042;

DT 09-NOV-1998 (first entry)

XX DE Human low density lipoprotein binding protein LBP-3.

KW Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;

KW receptor; human; atherosclerosis; diagnosis; therapy; vaccine.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 53..59

FT /note= "Claim 2"

XX WO9823282-A1.

XX 04-JUN-1998.

XX 26-NOV-1997; 97WO-US021857.

PR 27-NOV-1996; 96US-0031930P.

PR 03-JUN-1997; 97US-0048547P.

XX (BOST-) BOSTON HEART FOUND INC.

XX Lees AM, Lees RS, Law SW, Arjona AA;

XX WPI; 1998-322455/28.

DR N-PSDB; AAV32839.

XX Nucleic acid encoding low density lipoprotein binding proteins and

PT related vectors - transformed cells, proteins, and modulators of binding,

PT useful for treatment and diagnosis of atherosclerosis and for identifying

PT subjects at risk.

XX Claim 1; Fig 8; 47pp; English.

XX This polypeptide comprises novel human low density lipoprotein (LDL)

CC binding protein LBP-3 that is capable of binding both native and methyl

CC LDL. Its amino acid sequence was deduced from an isolated cDNA clone (see  
CC AAV32839). cDNA clones (see AAV32834-39) and encoded rabbit and human  
CC LBPs (see AAV49037-42) are claimed. An abnormality in an aspect of LBP  
CC metabolism or structure is diagnostic of a risk for atherosclerosis. The  
CC invention provides methods for determining if an animal is at risk for  
CC atherosclerosis (e.g. for prenatal screening); methods for treating  
CC atherosclerosis (including gene therapy) using e.g. LBP polypeptides to  
CC bind LDL and thereby prevent formation of atherosclerotic plaque; and  
CC methods for treating a cell having an abnormality in LBP structure or  
CC metabolism. Pharmaceutical and vaccine compositions are also provided, as  
CC well as recombinant vectors and host cells used to produce recombinant  
CC LBP

XX Sequence 530 AA;

Query Match 78.6%; Score 429; DB 2; Length 530;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAAPKPEGAQARTAQSGALRDVSEELS 76  
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAAPKPEGAQARTAQSGALRDVSEELS 60  
Qy 77 RQLEDILSTYCVDNNGGPGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136  
Db 61 RQLEDILSTYCVDNNGGPGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120  
Qy 137 KGDPTETIRQSDVGDHRRPQEKKAAGLKGKEITLLMQTLNTLSTPEKLAALCKKY 196  
Db 121 KGDPTETIRQSDVGDHRRPQEKKAAGLKGKEITLLMQTLNTLSTPEKLAALCKKY 180  
Qy 197 AELEEHNRNSQOMKLLQKKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQHRNRSK 256  
Db 181 AELEEHNRNSQOMKLLQKKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQHRNRSK 240  
Qy 257 EGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 316  
Db 241 EGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300  
Qy 317 ELREHIDKVPFKHQLQOOLVDKLAQQAQEMKLEAEERHOREKDFLLKEAVESQRCMEL 376  
Db 301 ELREHIDKVPFKHQLQOOLVDKLAQQAQEMKLEAEERHOREKDFLLKEAVESQRCMEL 360  
Qy 377 KQETHLKOOLALYTEKPEFQNTLSKSEVFTTFKQEMEKMTKKIKLEKETMYRSRW 436  
Db 361 KQETHLKOOLALYTEKPEFQNTLSKSEVFTTFKQEMEKMTKKIKLEKETMYRSRW 420  
Qy 437 ESSNKALLEMAEKTVRDKELEGQVKIQRIKLCRALQTERNDLNKRVQDLSAGQGSGL 496  
Db 421 ESSNKALLEMAEKTVRDKELEGQVKIQRIKLCRALQTERNDLNKRVQDLSAGQGSGL 480  
Qy 497 TDSGPERRPEGPGAQAPSSPRVTAPCPYEGAPSTEASGQTGPQPTSARA 546  
Db 481 TDSGPERRPEGPGAQAPSSPRVTAPCPYEGAPSTEASGQTGPQPTSARA 530

RESULT 7  
AA82804  
ID AA82804 standard; protein; 530 AA.  
XX AA82804;  
AC AA82804;  
XX  
DT 12-NOV-2001 (first entry)  
XX  
DE Human low density lipoprotein binding protein 3 (LBP-3).  
XX  
KW Low density lipoprotein binding protein 3; LBP-3; LDL; human;  
KW atherosclerosis; antiarteriosclerotic; therapy; diagnosis; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200164874-A2.  
XX

PD 07-SBP-2001.

XX  
XX 28-FEB-2001; 2001WO-US006356.  
PF  
XX 02-MAR-2000; 2000US-00517849.  
PR  
XX 14-JUL-2000; 2000US-00616289.  
PR  
XX (BOST-) BOSTON HEART FOUND INC.

XX  
XX Lees AM, Lees RS, Law SW, Arjona AA;  
PI  
XX WPI; 2001-565505/63.  
XX  
DR N-PSDB; AAH26496.

XX  
XX New isolated low density lipoprotein binding polypeptide for treating,  
PT diagnosing and/or identifying therapeutic agents for atherosclerosis.  
XX  
XX Claim 13(h); Fig 8B; 143pp; English.

XX  
CC The present sequence is that of the N-terminal portion of novel human low  
CC density lipoprotein binding protein 3 (LBP-3). The amino acid sequence is  
CC deduced from an isolated partial cDNA clone (see AAH26494). A full-length  
CC sequence is given in AA82808. Human LBP-3 is an example of claimed LBP  
CC polypeptides of the invention that are capable of binding to native and  
CC methylated low density lipoproteins. Also claimed are biologically active  
CC fragments and analogues of LBPs, polynucleotides encoding LBPs, as well  
CC as expression vectors, cells and methods of producing the LBPs. Methods  
CC for determining if an animal is at risk for atherosclerosis, and methods for  
CC evaluating an agent for use in treating atherosclerosis, and methods for  
CC treating a cell having an abnormality in structure or metabolism of LBP  
CC are claimed. Pharmaceutical compositions comprising an LBP polypeptide or  
CC nucleic acid, and vaccine compositions, are also claimed  
XX  
XX Sequence 530 AA;

Query Match 78.6%; Score 429; DB 4; Length 530;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAAPKPEGAQARTAQSGALRDVSEELS 76  
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAAPKPEGAQARTAQSGALRDVSEELS 60  
Qy 77 RQLEDILSTYCVDNNGGPGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 136  
Db 61 RQLEDILSTYCVDNNGGPGEDGAGQEPAPEDAEKSRITYVARNGEPEPTPVVNGEKEPS 120  
Qy 137 KGDPTETIRQSDVGDHRRPQEKKAAGLKGKEITLLMQTLNTLSTPEKLAALCKKY 196  
Db 121 KGDPTETIRQSDVGDHRRPQEKKAAGLKGKEITLLMQTLNTLSTPEKLAALCKKY 180  
Qy 197 AELEEHNRNSQOMKLLQKKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQHRNRSK 256  
Db 181 AELEEHNRNSQOMKLLQKKQSQLVQEKDHLRGEHSKAVLARSKLESICRELQHRNRSK 240  
Qy 257 EGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 316  
Db 241 EGVQARAEEREEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEKLLIEQY 300  
Qy 317 ELREHIDKVPFKHQLQOOLVDKLAQQAQEMKLEAEERHOREKDFLLKEAVESQRCMEL 376  
Db 301 ELREHIDKVPFKHQLQOOLVDKLAQQAQEMKLEAEERHOREKDFLLKEAVESQRCMEL 360  
Qy 377 KQETHLKOOLALYTEKPEFQNTLSKSEVFTTFKQEMEKMTKKIKLEKETMYRSRW 436  
Db 361 KQETHLKOOLALYTEKPEFQNTLSKSEVFTTFKQEMEKMTKKIKLEKETMYRSRW 420  
Qy 437 ESSNKALLEMAEKTVRDKELEGQVKIQRIKLCRALQTERNDLNKRVQDLSAGQGSGL 496  
Db 421 ESSNKALLEMAEKTVRDKELEGQVKIQRIKLCRALQTERNDLNKRVQDLSAGQGSGL 480  
Qy 497 TDSGPERRPEGPGAQAPSSPRVTAPCPYEGAPSTEASGQTGPQPTSARA 546  
Db



481 TDGSPRRPGCAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTPSARA 530

Db

## RESULT 8

AB007223

ID ABO07223 standard; protein; 356 AA.

XX

AC

XX

DT

XX

DE

XX

KW

XX

KW

XX

KW

XX

KW

XX

OS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

AB007223 standard; protein; 356 AA.

AC ABO07223;

DT 13-AUG-2003 (first entry)

XX Human p53 modifying protein, SEQ ID 183.

XX Human, p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;

XX antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;

XX lung cancer; ovarian cancer; angiogenesis; cell cycle;

XX apoptotic disorder; cell proliferation disorder.

XX Homo sapiens.

XX W0200299122-A1.

XX 12-DEC-2002.

XX 03-JUN-2002; 2002WO-US017382.

XX 05-JUN-2001; 2001US-0296076P.

XX 10-OCT-2001; 2001US-0328605P.

XX 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-136859/15.

XX N-PSDB; ACD33448.

XX Identifying modulators of the p53 pathway for use in treating apoptotic

XX or cell proliferation disorders, comprises screening for agents that

XX modulate activity of a human ortholog of genes that modify the p53

XX pathway in Drosophila.

XX Example 2; Page 533-534; 678pp; English.

XX The invention relates to identifying (M1) a candidate p53 pathway

XX modulating agent, by contacting an assay system comprising a purified HM

XX polypeptide (human orthologue of genes that modify the p53 pathway in

XX Drosophila) or nucleic acid with a test agent under conditions, where but

XX for the presence of the test agent, the system provides a reference

XX activity, and detecting a test agent-biased activity of the assay system.

XX Also included are modulating (M2) a p53 pathway of a cell (comprising

XX contacting a cell defective in p53 function with a candidate modulator

XX that specifically binds to a HM polypeptide comprising an HM amino acid

XX sequence, where p53 function is restored), modulating (M3) a p53 pathway

XX in a mammalian cell (comprising contacting the cell with an agent that

XX specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)

XX a disease in a patient (comprising: (a) obtaining a biological sample

XX from the patient; (b) contacting the sample with a probe for HM

XX expression; (c) comparing the results with a control; and (d) determining

XX whether the comparison indicates a likelihood disease). (M1) is useful

XX for identifying modulators of the p53 pathway. A probe for HM expression

XX is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,

XX in a patient, where the cancer has greater than 25 % expression level.

XX Modulators identified by (M1) are useful in a variety of diagnostic and

XX therapeutic applications, where disease or disorder prognosis is related

XX to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell

XX proliferation disorders (e.g. cancer). Another two new methods (M2 and

XX M3) are useful for modulating the p53 pathway of a cell, thus restoring

XX the p53 function of the cell, so that the cell undergoes normal

XX proliferation or progression through the cell cycle. (M2) and (M3) are

XX also useful for treating defects in the p53 pathway such as angiogenic,

XX apoptotic or cell proliferation disorders. The present sequence

XX represents a human p53 pathway modifying protein

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

SQ

Sequence 356 AA;

Query Match

Best Local Similarity

Matches 356; Conservative

65.2%; Score 356; DB 6; Length 356;

Pred. No. 0;

Mismatches 0;

Indels 0; Gaps 0;

QY

191

ALCKKYAELLEHRNSQKQMLLQKQSQQLVQKDHLEHSGKAVLARSKLESLCRLQ

250

DB

1

ALCKKYAELLEHRNSQKQMLLQKQSQQLVQKDHLEHSGKAVLARSKLESLCRLQ

60

QY

251

HNRSLKRGVQPARIEEKKREKVTSHFQVTLNDIOLQMEQHNRNSKLEHNLAEKRL

310

DB

61

HNRSLKRGVQPARIEEKKREKVTSHFQVTLNDIOLQMEQHNRNSKLEHNLAEKRL

120

QY

311

KLIEQVELREEHIDKVFHKLQKQQLVDAKLAQQAQEMLEKAEERHQRKDFLLKEAVSQ

370

DB

121

KLIEQVELREEHIDKVFHKLQKQQLVDAKLAQQAQEMLEKAEERHQRKDFLLKEAVSQ

180

QY

371

RMCCLMKQOETHLKOOLALYTEKFEFPONTLSKSSEVFTTFKQEVKMTKKIKLEKETT

430

DB

181

RMCCLMKQOETHLKOOLALYTEKFEFPONTLSKSSEVFTTFKQEVKMTKKIKLEKETT

240

QY

431

MYRSRWSSNKKALLEWAEKTVRDKELGLQVKIORLEKLCALQTERNDLNKRVQDLSA

490

DB

241

MYRSRWSSNKKALLEWAEKTVRDKELGLQVKIORLEKLCALQTERNDLNKRVQDLSA

300

QY

491

GGQGSILTDGSPRRPPEGGAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTPSARA

546

DB

301

GGQGSILTDGSPRRPPEGGAQAPSSPRVTEAPCYPCAPSTEASGQTGPQPTPSARA

356

RESULT 9

AAM78757

ID AAM78757 standard; protein; 358 AA.

XX

AC

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Human protein SEQ ID NO 1419.

Human; cytokine; cell proliferation; cell differentiation; gene therapy;

vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

tissue growth factor; immunomodulatory; cancer; leukaemia;

nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX W0200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 27-APR-2000; 2000US-00560875.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00620325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejman T, Goodrich R;

XX WPI; 2001-476283/51.

XX N-PSDB; AAK51890.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
in diagnosis and gene therapy.  
XX  
XX  
PS Claim 20; Page 3681-3682; 6221pp; English.  
XX  
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapies. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
CC sequence listing were missing at the time of publication  
XX  
XX Sequence 358 AA;  
SQ  
Query Match 63.6%; Score 347; DB 4; Length 358;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 200 LEEHNSQKQMLKQKQSQVQEKDHLRGHSHKAVLARSKLBSLCRELOHNSLKEEG 259  
Db 12 LEEHNSQKQMLKQKQSQVQEKDHLRGHSHKAVLARSKLBSLCRELOHNSLKEEG 71  
Qy 260 VQARAEERKKEVTSHTFQVTLNDIQLOMEQHNSKLRQENNELAERLKLIEQVELR 319  
Db 72 VQARAEERKKEVTSHTFQVTLNDIQLOMEQHNSKLRQENNELAERLKLIEQVELR 131  
Qy 320 EEHIDKVFKHDIQQLVDAKQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELMKQ 379  
Db 132 EEHIDKVFKHDIQQLVDAKQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELMKQ 191  
Qy 380 ETHLKQQLALYTEKPEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 439  
Db 192 ETHLKQQLALYTEKPEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 251  
Qy 440 NKALLEMAEKTVDKLEGLQVKIORLEKLCALQTERNDLNKRVODLSAGGQGSITDS 499  
Db 252 NKALLEMAEKTVDKLEGLQVKIORLEKLCALQTERNDLNKRVODLSAGGQGSITDS 311  
Qy 500 GPERRPEGGAQAPSSPRVTBAPCYGAPSTEASGQTGPQPTTSARA 546  
Db 312 GPERRPEGGAQAPSSPRVTBAPCYGAPSTEASGQTGPQPTTSARA 358  
RESULT 10  
ABG00839  
ID ABG00839 standard; protein; 1749 AA.  
XX  
XX ACG00839;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #830.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
XX  
XX 31-MAR-2000; 2000US-00540217.  
XX  
XX 23-AUG-2000; 2000US-00649167.  
XX

XX (HYSE-) HYSEQ INC.  
PA Dmanac RT, Liu C, Tang YT;  
XX  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS65026.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 20; SEQ ID NO 31198; 103pp; English.  
PS  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1749 AA;  
Query Match 41.8%; Score 228; DB 4; Length 1749;  
Best Local Similarity 100.0%; Pred. No. 1.1e-211;  
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 257 EGVQARAEERKKEVTSHTFQVTLNDIQLOMEQHNSKLRQENNELAERLKLIEQY 316  
Db 438 EGVQARAEERKKEVTSHTFQVTLNDIQLOMEQHNSKLRQENNELAERLKLIEQY 497  
Qy 317 ELREEHIDKVFKHDIQQLVDAKQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 376  
Db 498 ELREEHIDKVFKHDIQQLVDAKQQAQEMLKEAEERHOREKDFLLKEAVESQRMCELM 557  
Qy 377 KQOETHLKQQLALYTEKPEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 436  
Db 558 KQOETHLKQQLALYTEKPEFQNTLSKSEVFTTFQEMEKMTKKIKLEKETTMYRSRW 617  
Qy 437 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCALQTERNDLNK 484  
Db 618 ESSNKALLEMAEKTVDKLEGLQVKIORLEKLCALQTERNDLNK 665  
RESULT 11  
AAW49039  
ID AAW49039 standard; protein; 557 AA.  
XX  
XX AAW49039;  
XX  
XX 09-NOV-1998 (first entry)  
XX  
XX Rabbit low density lipoprotein binding protein LBP-3.  
XX  
XX Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;  
XX receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.  
XX



```

OS Homo sapiens.
XX WO2000073801-A2.
XX PD 07-DEC-2000.
XX PF 26-MAY-2000; 2000WO-US014749.
XX PR 28-MAY-1999; 99US-0136526P.
XX PR 10-SEP-1999; 99US-0153454P.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Obata Y;
XX DR WP1; 2001-025274/03.
XX PT Nucleic acids encoding breast, gastric and prostate cancer associated
XX PT antigen precursors, useful for diagnosing and treating a condition
XX PT characterized by expression of an abnormal amount of a protein, e.g.
XX PT cancer.
XX PS Example 1; Page 486-487; 799pp; English.
XX CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
XX CC represent nucleotide sequences encoding human breast, gastric and
XX CC prostate cancer associated antigen precursors (CAAP), respectively.
XX CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
XX CC represent human breast, gastric and prostate CAAP protein sequence
XX CC respectively. CAAPs have cytostatic activity and can be used in the
XX CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
XX CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
XX CC condition characterised by expression of an abnormal amount of a protein,
XX CC e.g. cancer.
XX SQ Sequence 204 AA;

Query Match 3.5%; Score 19; DB 4; Length 204;
Best Local Similarity 100.0%; Pred. No. 1.3e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LNTLSTPEEKLAALCKKYA 197
Db 1 LNTLSTPEEKLAALCKKYA 19
|||||
|||||

RESULT 14
AAW37883
XX ID AAW37883 standard; protein; 386 AA.
XX AC AAW37883;
XX DT 28-AUG-1998 (first entry)
XX DE BRCA1 modulator protein 091-132Q20.
XX KW BRCA1 modulator protein; 091-132Q20; breast cancer antigen 1;
XX KW tumour suppressor protein; diagnosis; therapy; human.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 124..143
XX FT /note= "leucine zipper motif"
XX PN WO9810066-A1.
XX PD 12-MAR-1998.
XX PF 06-AUG-1997; 97WO-US013944.
XX PR 04-SEP-1996; 96US-0025601P.
XX PT

(ONYX-) ONYX PHARM INC.
Rubinfeld B, Polakis P, Ligenfelter C, Vuong TT;
WPI; 1998-193616/17.
N-PSDB; AAV29064.
Breast cancer antigen 1 modulator protein - useful for diagnosing
diseases involving unwanted cell growth, e.g. breast cancer, and for
producing therapeutics for treatment of such diseases.
Example 1; Fig 3; 73pp; English.
This polypeptide comprises a 46 kDa BRCA1 modulator protein that binds to
the tumour suppressor gene product BRCA1, and which is characterised by a
leucine zipper motif. Its amino acid sequence was deduced from the
nucleotide sequence of a cDNA clone (see AAV29064), designated 091-132Q20
(ATCC 98143), isolated from a HeLa cell cDNA library using a yeast two-
hybrid assay. 3 cDNA clones (see also AAV29062 and AAV29063) coding for
BRCA1 modulator proteins (see AAW37881-83) have been characterised.
Vectors and host cells comprising the isolated nucleic acid sequences are
claimed, as well as a process for producing BRCA1 modulator protein by
culturing these host cells. BRCA1 modulator proteins and nucleic acids
can be used to diagnose diseases involving unwanted cell growth, e.g.
breast cancer, and to identify compounds that alter BRCA1 interaction
with BRCA1 modulators for the treatment of such diseases
XX SQ Sequence 386 AA;

Query Match 3.5%; Score 19; DB 2; Length 386;
Best Local Similarity 100.0%; Pred. No. 2.2e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 179 LNTLSTPEEKLAALCKKYA 197
Db 135 LNTLSTPEEKLAALCKKYA 153
|||||
|||||

RESULT 15
AAV30151
XX ID AAV30151 standard; protein; 386 AA.
XX AC AAV30151;
XX DT 27-OCT-1999 (first entry)
XX DE Amino acid sequence of a BRCA1 modulator protein.
XX KW Modulator protein; BRCA1; tumour suppressor protein; breast cancer;
XX KW ovarian cancer; cell growth; cell proliferation.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Region 125..143
XX FT /note= "leucine zipper motif"
XX PN US5948643-A.
XX PD 07-SEP-1999.
XX PF 13-AUG-1997; 97US-00968751.
XX PR 13-AUG-1997; 97US-00968751.
XX PA (ONYX-) ONYX PHARM INC.
XX PI Rubinfeld B, Ligenfelter C, Vuong TT, Polakis PG;
XX WPI; 1999-517952/43.
XX DR N-PSDB; AAX86756.
XX PT Modulator proteins that bind to and modulate the activity of the BRCA1

```

Search completed: June 8, 2004, 16:37:53  
Job time : 62 secs

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:36:07 ; Search time 21 Seconds  
(without alignments)  
2500.980 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 546

Sequence: 1 MNQDKKNGAAGQSNPKSP.....APSTEASGQTGPQPTSPARA 546

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR 78.\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	3.1	676	2	JC7222	77K muscle-derived
2	9	1.6	594	2	S50611	hypothetical prote
3	9	1.6	1402	2	I46707	translation initia
4	8	1.5	162	2	T29728	hypothetical prote
5	8	1.5	211	2	T32354	hypothetical prote
6	8	1.5	224	2	T52558	translation elonga
7	8	1.5	229	2	E87236	probable membrane
8	8	1.5	260	2	S71315	deoxyguanosine kin
9	8	1.5	281	2	C88638	protein P58f6.1 [i
10	8	1.5	312	2	T08985	hypothetical prote
11	8	1.5	450	2	E71909	hypothetical prote
12	8	1.5	467	2	D88710	protein C43G2.1 [i
13	8	1.5	500	2	H97769	hypothetical prote
14	8	1.5	525	2	H75514	glutamyl-tRNA synt
15	8	1.5	531	2	A55877	caldesmon, non-mus
16	8	1.5	535	2	C36811	hypothetical prote
17	8	1.5	535	2	A37994	RPI protein - salm
18	8	1.5	568	2	E90364	hypothetical prote
19	8	1.5	594	2	S62141	transcription init
20	8	1.5	758	2	S65169	hypothetical prote
21	8	1.5	761	2	T32193	hypothetical prote
22	8	1.5	928	2	T52292	endopeptidase Clp
23	8	1.5	952	2	T52456	endopeptidase Clp
24	8	1.5	952	2	T49283	AtClpC - Arabidops
25	8	1.5	1100	2	AE3243	conjugal transfer
26	8	1.5	1101	2	T03419	traA protein - Agr
27	8	1.5	1902	2	C97702	cell surface antig
28	8	1.5	2052	2	T18519	myosin X - bovine
29	7	1.3	23	2	S45032	homeotic protein S

ALIGNMENTS

RESULT 1

JC7222

77K muscle-derived protein - chicken

C:Species: Gallus gallus (chicken)

C>Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000

C/Accession: JC7222

R/Dyeda, A.; Fukui, I.; Fujimori, K.; Kiyosue, K.; Nishimune, H.; Kasai, M.; Taguchi, T

Biochem. Biophys. Res. Commun. 269, 564-569, 2000

A:Title: MDP77: A novel neurite-outgrowth-promoting protein predominantly expressed in

A:Reference number: JC7222; PMID:20175243; PMID:10708594

A/Accession: JC7222

A/Molecule type: mRNA

A/Residues: 1-676 <UYE>

A/Cross-references: GB:D89999; NID:G7619883; PIDN:BAA94755.1; PID:G7619884

A/Experimental source: crus muscle

C/Comment: This protein, a glycoprotein and a neurite-outgrowth-promoting protein, is i

C/Keywords: coiled coil; glycoprotein; leucine zipper; muscle

Query Match 3.1%; Score 17; DB 2; Length 676;

Best Local Similarity 100.0%; Pred. No. 2.1e-08;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 LARSKLESICRLQRHN 252

Db 207 LARSKLESICRLQRHN 223

RESULT 2

S50611

hypothetical protein YER108c - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C>Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 23-Mar-2001

C/Accession: S50611

R/Dietrich, F.S.

submitted to the EMBL Data Library, December 1994

A/Description: The sequence of S. cerevisiae cosmids 9781, 8198, 9115, 9981, and lambda

A/Reference number: S50437

A/Accession: S50611

A/Molecule type: DNA

A/Residues: 1-594 <PIE>

A/Cross-references: EMBL:U18916; NID:G1384128; PIDN:AC03206.1; PID:G603347; MIPS:YER10

C/Genetics:

A/Gene: SGD:FLO8

A/Cross-references: SGD:S0000910; MIPS:YER108c

A/Map position: 5R

Query Match 1.6%; Score 9; DB 2; Length 594;

Best Local Similarity 100.0%; Pred. No. 2.3;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 NTLSTPEEK 188

Db 466 NTLSTPEK 474

RESULT 3  
I46707  
translation initiation factor eIF4-gamma - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Feb-1997 #sequence\_revision 14-Feb-1997 #text\_change 31-Mar-2000  
C:Accession: I46707  
R:Yan, R.; He, W.; Rhoads, R.E.  
J. Biol. Chem. 268, 19200-19203, 1993  
A:Title: Mapping the cleavage site in protein synthesis initiation factor eIF-4 gamma of  
A:Reference number: I46707; PMID:8396129  
A:Accession: I46707  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1402 <YAN>  
A:Cross-references: GB:L22090; NID:G404774; PIDN:AAA31242.1; PID:G404775

Query Match 1.6%; Score 9; DB 2; Length 1402;  
Best Local Similarity 100.0%; Pred. No. 5.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 72 SEELSRQLE 80  
|||||  
Db 1243 SEELSRQLE 1251

RESULT 4  
T29728  
hypothetical protein K03B4.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29728  
R:Du, Z.; Le, T.T.; Kemp, K.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid K03B4.  
A:Reference number: Z20673  
A:Accession: T29728  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-162 <DUZ>  
A:Cross-references: EMBL:U55370; PIDN:AAA97999.1; GSPDB:GN00023; CESP:K03B4.7  
A:Experimental source: strain Bristol N2; clone K03B4  
C:Genetics:  
A:Gene: CESP:K03B4.7  
A:Map position: 5  
A:Introns: 58/3; 105/3

Query Match 1.5%; Score 8; DB 2; Length 162;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 AERLKKLI 313  
|||||  
Db 89 AERLKKLI 96

RESULT 5  
T32354  
hypothetical protein C08E3.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32354  
R:Miller, N.; Kramer, J.; Keppeler, D.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid C08E3.  
A:Reference number: Z21155  
A:Accession: T32354  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-211 <ML>  
A:Cross-references: EMBL:AF025457; PIDN:ABF70965.1; GSPDB:GN00020; CESP:C08E3.4

A:Experimental source: strain Bristol N2; clone C08E3  
C:Genetics:  
A:Gene: CESP:C08E3.4  
A:Map position: 2  
A:Introns: 36/3; 149/3

Query Match 1.5%; Score 8; DB 2; Length 211;  
Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 AERLKKLI 313  
|||||  
Db 81 AERLKKLI 88

RESULT 6  
T52558  
translation elongation factor eEF1Balpha (clone 2) [validated] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 24-Oct-2000 #sequence\_revision 24-Oct-2000 #text\_change 07-Jul-2003  
C:Accession: T52558  
R:Hericourt, P.; Jupin, I.  
FEBS Lett. 464, 148-152, 1999  
A:Title: Molecular cloning and characterization of the Arabidopsis thaliana alpha-subu  
A:Reference number: Z26114  
A:Accession: T52558  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-224 <HER>  
A:Cross-references: EMBL:AJ249597; PIDN:CAB64730.1  
C:Genetics:  
A:Gene: eEF1Balpha2  
C:Function:  
A:Description: involved in translation elongation; able to complement a mutant yeast s  
C:Superfamily: translation elongation factor eEF-1 beta chain; translation elongation f

Query Match 1.5%; Score 8; DB 2; Length 224;  
Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 AAPAVEAE 44  
|||||  
Db 84 AAPAVEAE 91

RESULT 7  
E87236  
probable membrane protein. [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: E87236  
R:Cole, S.T.; Eiglmeier, K.; Parthill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; F  
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd  
eam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S  
A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: A86909; PMID:21128732; PMID:11234002  
A:Accession: E87236  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-229 <SFO>  
A:Cross-references: GB:AL450380; NID:G13093841; PIDN:CAC32147.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: ML2615  
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0200

Query Match 1.5%; Score 8; DB 2; Length 229;  
Best Local Similarity 100.0%; Pred. No. 9.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 PSQAAPAV 41  
|||||

Db 200 PQAPAV 207

RESULT 8  
S71315  
deoxyguanosine kinase (EC 2.7.1.113) precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 20-Jun-2000  
C:Accession: S71315; S78432  
R:Wang, L.; Hellman, U.; Eriksson, S.  
FEBS Lett. 390, 39-43, 1996  
A:Title: Cloning and expression of human mitochondrial deoxyguanosine kinase cDNA.  
A:Reference number: S71315; MUID:96314545; PMID:8706825  
A:Accession: S71315  
A:Molecule type: mRNA  
A:Residues: 1-260 <MAN>  
A:Cross-references: EMBL:X97386; PIDN:CAA66054.1  
A:Experimental source: tissue brain  
R:Wang, L.; Hellman, U.; Eriksson, S.  
submitted to the EMBL Data Library, April 1996  
A:Description: Cloning and expression of human deoxyguanosine kinase cDNA.  
A:Reference number: S78432  
A:Accession: S78432  
A:Molecule type: mRNA  
A:Residues: 1-18; R', 20-260 <MAN>  
A:Cross-references: EMBL:X97386; PIDN:CAA66054.1  
A:Experimental source: brain  
C:Genetics: dgk  
A:Gene: dgk  
A:Genome: nuclear  
C:Superfamily: human deoxycytidine kinase  
C:Keywords: mitochondrion; phosphotransferase  
F:1-22/Domain: transit peptide (mitochondrion) #status predicted <TNP>  
F:23-260/Product: deoxyguanosine kinase #status predicted <MAT>  
F:125-132/Region: DRS motif  
F:185-191/Region: arginine-rich

Query Match 1.5%; Score 8; DB 2; Length 260;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 QRAREEE 268  
| | | | |  
Db 188 QRAREEE 195

RESULT 9  
C88638  
protein F58F6.1 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 15-Sep-2003  
C:Accession: C88638  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: C88638  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-281 <STO>  
A:Cross-references: GB:chr\_IV; PIDN:AAB88358.1; PID:G2662599; GSPDB:GN00022; CRSP:F58F6.  
C:Genetics:  
A:Gene: F58F6.1  
A:Map position: 4

Query Match 1.5%; Score 8; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 GPGEDGA 100  
| | | | |

Db 241 GPGEDGA 248

RESULT 10  
T08985  
hypotheical protein F6G3.120 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
C:Accession: T08985  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.P.  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16520  
A:Accession: T08985  
A:Molecule type: DNA  
A:Residues: 1-312 <BEV>  
A:Cross-references: EMBL:AL078464; GSPDB:GN00062; ATSP:F6G3.120  
A:Experimental source: cultivar Columbia; BAC clone F6G3  
C:Genetics:  
A:Gene: ATSP:F6G3.120  
A:Map position: 4  
A:Introns: 55/1; 94/3; 115/3; 151/3; 168/3; 189/3; 241/3; 264/3; 285/2

Query Match 1.5%; Score 8; DB 2; Length 312;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 342 QQAQEMLK 349  
| | | | |  
Db 240 QQAQEMLK 247

RESULT 11  
E71909  
hypotheical protein jhp0612 - Helicobacter pylori (strain J99)  
C:Species: Helicobacter pylori  
A:Variety: strain J99  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 08-Oct-1999  
C:Accession: E71909  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.B.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: E71909  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-450 <ARN>  
A:Cross-references: GB:AB001493; GB:AB001439; NID:94155161; PIDN:AAD06193.1; PID:941551  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0612

Query Match 1.5%; Score 8; DB 2; Length 450;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 LEDILSTY 86  
| | | | |  
Db 283 LEDILSTY 290

RESULT 12  
D88710  
protein C43G2.1 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: D88710  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_el  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and



A/Accession: D88710  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-467 <STO>  
 A/Cross-references: GB:chr\_IV; PIDN:AA09107.1; PID:gl512756; GSPDB:GN00022; CESP:C43G2  
 C/Genetics:  
 A/Gene: C43G2.1  
 A/Map position: 4

Query Match 1.5%; Score 8; DB 2; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 189 LAALCKXY 196  
 |||||  
 Db 22 LAALCKXY 29

## RESULT 13

H97769  
 Hypothetical protein murD [imported] - Rickettsia conorii (strain Malish 7)  
 C/Species: Rickettsia conorii  
 C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 22-Oct-2001  
 C/Accession: H97769  
 R/Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rickettsia conorii  
 Science 293, 2093-2096, 2001  
 A/Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
 A/Reference number: A97700; MUID:21442074; PMID:11557893  
 A/Accession: H97769  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-500 <KUR>  
 A/Cross-references: GB:AR006914; PIDN:AA03098.1; PID:gl5619640; GSPDB:GN00173  
 C/Genetics:  
 A/Gene: murD  
 C/Superfamily: UDP-N-acetylmuramate-alanine ligase

Query Match 1.5%; Score 8; DB 2; Length 500;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 QGEPAPE 108  
 |||||  
 Db 278 QGEPAPE 285

## RESULT 14

H75514  
 Glutamy1-tRNA synthetase - Deinococcus radiodurans (strain R1)  
 C/Species: Deinococcus radiodurans  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C/Accession: H75514  
 R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A/Reference number: A75250; MUID:20036896; PMID:10567266  
 A/Accession: H75514  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-525 <WHI>  
 A/Cross-references: GB:AR001907; GB:AE000513; NID:ge458162; PIDN:AAFI0063.1; PID:ge45817  
 C/Genetics:  
 A/Experimental source: strain R1  
 A/Gene: DR0485  
 A/Map position: 1  
 C/Superfamily: glutamate-tRNA ligase; glutamine-tRNA ligase homology

Query Match 1.5%; Score 8; DB 2; Length 525;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 512 APSSPRVT 519  
 |||||  
 Db 45 APSSPRVT 52

## RESULT 15

A55887  
 caldesmon, non-muscle - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 13-Aug-1999  
 C/Accession: A55887  
 R/Yamashita, S.; Yamakita, Y.; Yoshida, K.; Takiguchi, K.; Matsumura, F.  
 J. Biol. Chem. 270, 4023-4030, 1995  
 A/Title: Characterization of the COOH terminus of non-muscle caldesmon mutants lacking  
 A/Reference number: A55887; MUID:95181370; PMID:7876150  
 A/Accession: A55887  
 A/Status: preliminary; nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 1-531 <YAM>  
 A/Cross-references: GB:U18419; NID:ge222966; PIDN:AA068521.1; PID:ge222967  
 A/Note: authors translated the codon GCC for residue 68 as Val  
 C/Superfamily: caldesmon  
 C/Keywords: phosphoprotein

Query Match 1.5%; Score 8; DB 2; Length 531;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 263 AREREKR 270  
 |||||  
 Db 334 AREREKR 341

Search completed: June 8, 2004, 16:39:54  
 Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:34:26 ; Search time 17 Seconds  
(without alignments)  
1672.371 Million cell updates/sec

Title: US-10-023-529-44  
Perfect score: 546  
Sequence: 1 MKNQDKKGAQKSPKSP.....APSTRASQGTGPQEPSTAPA 546

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	259	47.4	259	YL14 HUMAN	P40222 homo sapien
2	9	1.6	799	F108 YEAST	Q4068 saccharomyc
3	9	1.6	1070	CARB METAC	Q8tny4 methanosarc
4	9	1.6	1073	CARB METWA	P58944 methanosarc
5	9	1.6	1402	IP4G RABIT	P41110 oryctolagus
6	8	1.5	483	SYE DEIRA	Q9rx30 deinococcus
7	8	1.5	500	MURD RICON	Q92i60 rickettsia
8	8	1.5	531	CALD RAT	Q62736 rattus norv
9	8	1.5	535	VG50 HSVSA	Q01012 herpesvirus
10	8	1.5	594	TFCS YEAST	P46678 saccharomyc
11	8	1.5	758	YPS8 YEAST	Q99299 saccharomyc
12	8	1.5	1070	RPOB LOTJA	Q9bb99 lotus japon
13	8	1.5	1072	CARB THERTN	Q8rbx0 thermoanaer
14	8	1.5	1100	TRAA AGRTS	Q44349 agrobacteri
15	8	1.5	2052	MT10 EOVIN	P79114 bos taurus
16	7	1.3	105	RK21 ODOSI	P49557 oodentelia s
17	7	1.3	121	INSC ECOLI	P19776 escherichia
18	7	1.3	121	INSC SHIPL	P59444 shigella fl
19	7	1.3	127	CYTC RAT	P14841 rattus norv
20	7	1.3	138	NE13 PARHU	Q40905 parietaria
21	7	1.3	139	NU11 PARUJ	P43217 parietaria
22	7	1.3	176	NU12 PARUJ	Q04404 parietaria
23	7	1.3	186	RFP WOLSU	Q7mad8 wolfinella s
24	7	1.3	187	ISP2 VIBPA	Q87na5 vibrio para
25	7	1.3	187	ISP2 VIBU	P59366 vibrio vuln
26	7	1.3	193	LOLB NEIMA	P57023 neisseria m
27	7	1.3	193	LOLB NEIMS	P57024 neisseria m
28	7	1.3	195	AAANT HDVDJ	P29995 hepatitis d
29	7	1.3	195	AAANT HDVLI	P29833 hepatitis d
30	7	1.3	195	AAANT HDVMI	P25881 hepatitis d
31	7	1.3	195	AAANT HDVNI	P25882 hepatitis d
32	7	1.3	199	VATE PYRAB	Q9uxu4 pyrococcus
33	7	1.3	207	TRPF_STAB	Q8csn5 staphylococ

34	7	1.3	211	1	YA03 ARCFU	O29259 archaeoglob
35	7	1.3	214	1	AAANT HDVAM	P25989 hepatitis d
36	7	1.3	214	1	AAANT HDVNA	P25880 hepatitis d
37	7	1.3	218	1	SGAH MYCPN	P75293 mycoplasma
38	7	1.3	220	1	R929 HUMAN	O95707 homo sapien
39	7	1.3	221	1	YNOC ECOLI	P76114 escherichia
40	7	1.3	225	1	TIHY ARATH	Q9sw33 arabidopsis
41	7	1.3	227	1	PYRH ASRPE	Q9y740 aeropyrum p
42	7	1.3	229	1	REGQ BP82	P13870 bacterioph
43	7	1.3	248	1	MYFO HUMAN	P25189 homo sapien
44	7	1.3	277	1	DGK HUMAN	Q16854 homo sapien
45	7	1.3	297	1	PERE ECOLI	Q57083 escherichia

ALIGNMENTS

RESULT 1  
YL14 HUMAN  
ID YL14 HUMAN STANDARD; PRT; 259 AA.  
AC P40222; Q8N2Y3;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical protein initially thought to be identical with  
DE interleukin-14 (IL-14) (High molecular weight B-cell growth factor)  
DE (HMW-BCGP).  
DE IL14.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93317675; PubMed=8327514;  
RA Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,  
RA Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;  
RT "Identification of a cDNA for a human high-molecular-weight B-cell  
RT growth factor."  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6330-6334(1993).  
RN [2]  
RP ERRATUM.  
RX MEDLINE=96353961; PubMed=8755619;  
RA Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,  
RA Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;  
RL Proc. Natl. Acad. Sci. U.S.A. 93:8154-8154(1996).  
RN [3]  
RP SEQUENCE OF 130-259 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Pahay J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shcherbenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- CAUTION: The clone described in Ref.1 was initially thought to  
CC code for a high molecular weight interleukin (IL-14). The  
CC translation of this cDNA does not predict the open reading frame

CC for the 60 kDa protein described in Ref.1. A reading frame on the  
CC plus strand predicts a 7.7 kDa protein. The longest open reading  
CC frame (shown here) is on the opposite strand and predicts a 36.4  
CC kDa protein. The relationship of this sequence to IL-14, if any,  
CC is uncertain.

CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; L15344; -; NOT ANNOTATED CDS.  
CC EMBL; BC029686; AAH29686.1; -;  
CC PIR; A48203; A48203.  
CC Genew; HGNC:5976; IL14.  
CC MIM; 147684; -;  
CC GO; GO:0005576; C:extracellular; NAS.  
CC GO; GO:0030372; F:high molecular weight B-cell growth factor . . .; NAS.  
CC GO; GO:0008283; P:cell proliferation; NAS.  
CC KW Hypothetical protein.  
CC SEQUENCE 259 AA; 30107 MW; 8AB8D1AA6891357C CRC64;

Query Match 47.4%; Score 259; DB 1; Length 259;  
Best Local Similarity 100.0%; Pred. No. 1.1e-249; Indels 0; Gaps 0;  
Matches 259; Conservative 0; Mismatches 0;

QY 288 MEQHNRNKLQENMELAEKLEIQYELREHIDKVFVKHDLQQLVDKLAQAQEM 347  
DB 1 MEQHNRNKLQENMELAEKLEIQYELREHIDKVFVKHDLQQLVDKLAQAQEM 60  
QY 348 LKAEERHQREDFLLKEAVESQRCMLKQOETHLKOALALYTKPEFQNTLSKSEV 407  
DB 61 LKAEERHQREDFLLKEAVESQRCMLKQOETHLKOALALYTKPEFQNTLSKSEV 120  
QY 408 FTTFQKMEKMTKKIKLEKETTYRSWSSNKALLEAEKTYRDKELSLQVKIQRL 467  
DB 121 FTTFQKMEKMTKKIKLEKETTYRSWSSNKALLEAEKTYRDKELSLQVKIQRL 180  
QY 468 EKLCLALQTERNDLNKRVODLSAGQGGSLTDSGPPRRPGEPGAQAPSPRVTEAPCYGA 527  
DB 181 EKLCLALQTERNDLNKRVODLSAGQGGSLTDSGPPRRPGEPGAQAPSPRVTEAPCYGA 240  
QY 528 PSTEASGQTGPQEPSARA 546  
DB 241 PSTEASGQTGPQEPSARA 259

RESULT 2  
FLO8 YEAST  
ID FLO8 YEAST STANDARD; PRT; 799 AA.  
AC P40068; P40067; Q05751;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Transcriptional activator FLO8 (PDH5 protein).  
GN FLO8 OR PDH5 OR YER109C/YER108C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=Sigma 1278B;  
RA Liu H., Styles C.A., Fink G.;  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RW [2]  
RC SEQUENCE OF 71-799 FROM N.A.  
RC STRAIN=ATCC 60715;  
RX MEDLINE=96335146; PubMed=8757402;  
RA Kobayashi O., Suda H., Ohtani T., Some H.;

RT "Molecular cloning and analysis of the dominant flocculation gene  
RT FLO8 from Saccharomyces cerevisiae.";  
RL Mol. Gen. Genet. 251:707-715(1996).  
RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RX MEDLINE=9733264; PubMed=9169868;  
RA Dietrich F.S., Mulligan J.T., Hennessy K.M., Yelton M.A., Allen E.,  
RA Araujo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,  
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,  
RA Hunnicke-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,  
RA Lin D., Mossdale D., Nakahara K., Namath A., Norgren R., Oefner P.,  
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,  
RA Smith V., Taylor P., Wei Y., Botstein D., Davis R.W.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V";  
RL Nature 387:78-81(1997).

CC -!- FUNCTION: Required for diploid filamentous growth, haploid  
CC invasive growth and flocculation. Putative transcriptional  
CC activator of FLO1.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: Contains 1 Lish domain.  
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a  
CC frameshift.

CC -----  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; U51431; AAC49522.1; -;  
CC EMBL; D83713; BAA12076.1; -;  
CC EMBL; U18916; AAC03207.1; ALT\_FRAME.  
CC EMBL; U18916; AAC03206.1; ALT\_FRAME.  
CC GenOnline; 139188; -;  
CC TRANSFAC; T03311; -;  
CC SGD; S0000911; FLO8.  
CC GO; GO:0005737; C:cytoplasm; IDA.  
CC GO; GO:0005634; C:nucleus; IDA.  
CC InterPro; IPR006594; Lish.  
CC SMART; SM00667; Lish; 1.  
CC PROSITE; PS00896; LISH; 1.  
CC Transcription regulation; Activator; Nuclear protein.  
CC DOMAIN 73 105  
CC POLY-GLN.  
CC FT CONFLICT 112 112 V -> I (IN REF. 2).  
CC FT CONFLICT 115 115 P -> S (IN REF. 2).  
CC FT CONFLICT 383 383 C -> G (IN REF. 2).  
CC FT CONFLICT 441 441 A -> T (IN REF. 2).  
CC FT CONFLICT 447 447 A -> V (IN REF. 2).  
CC FT CONFLICT 598 598 R -> P (IN REF. 2).  
CC SEQUENCE 799 AA; 86834 MW; 91AAE10D8E586DAF CRC64;

Query Match 1.6%; Score 9; DB 1; Length 799;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 NTLSTPEEK 188  
DB 671 NTLSTPEEK 679  
RESULT 3  
CARB METAC STANDARD; PRT; 1070 AA.  
ID CARB METAC  
AC Q8TNY4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-  
DE phosphate synthetase ammonia chain).

GY CARB OR MA2143.  
 OS Methanosarcina acetivorans.  
 CC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 CC Methanosarcinaceae; Methanosarcina.  
 OK NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238;  
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,  
 RA Fitzhugh W.E., Calvo S., Engels R., Smirnov S., Atcoor D., Brown A.,  
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,  
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,  
 RA Zimmer A., Barber R.D., Cann I., Graham D.B., Krzycki J.A., Guss A.M.,  
 RA Hedgerich R., Ingram-Smith C., Kuetner H.C., Kzycki J.A.,  
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,  
 RA Springer T.A., Umayal L.A., White O., White R.H., de Macario E.C.,  
 RA Perry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,  
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,  
 RA Metcalf W.W., Birren B.  
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic  
 RT and physiological diversity."  
 RL Genome Res. 12:532-542(2002).  
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +  
 CC phosphate + L-glutamate + carbamoyl phosphate.  
 CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).  
 CC -!- PATHWAY: Arginine biosynthesis.  
 CC -!- SUBUNIT: Pyrimidine biosynthesis; first step.  
 CC promotes the hydrolysis of glutamine to ammonia, which is used by  
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to the carb family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE010901; AM05541.1; --  
 CC HAMAP; MF\_01210; 1;  
 CC InterPro; IPR006275; CarA\_L\_glu.  
 CC InterPro; IPR005483; CPase\_L.  
 CC InterPro; IPR005479; CPase\_L\_D2.  
 CC InterPro; IPR005480; CPase\_L\_D3.  
 CC InterPro; IPR005481; CPase\_L\_N.  
 CC InterPro; IPR004362; MGS like.  
 CC Pfam; PF00289; CPase\_L\_chain; 2.  
 CC Pfam; PF02786; CPase\_L\_D2; 2.  
 CC Pfam; PF02787; CPase\_L\_D3; 1.  
 CC Pfam; PF02142; MGS; 1.  
 CC PRINTS; PR00098; CPASE.  
 CC TIGRFAMs; TIGR01369; CPaseII\_lrg; 1.  
 CC PROSITE; PS00865; CPASE\_1; 1.  
 CC PROSITE; PS00867; CPASE\_2; 1.  
 CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;  
 KW ATP-binding; Manganese; Complete proteome.  
 FT DOMAIN 1 399  
 FT DOMAIN 400 540 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.  
 FT DOMAIN 541 931 OLIGOMERIZATION DOMAIN.  
 FT DOMAIN 932 1070 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.  
 FT REPEAT 1 540 ALLOSTERIC DOMAIN.  
 FT REPEAT 541 1070  
 FT NP BIND 153 210 ATP (POTENTIAL).  
 FT NP BIND 284 350 ATP (POTENTIAL).  
 FT METAL 284 350 MANGANESE 1 (BY SIMILARITY).  
 FT METAL 296 296 MANGANESE 1 AND 2 (BY SIMILARITY).  
 FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).  
 FT METAL 822 822 MANGANESE 3 (BY SIMILARITY).  
 FT METAL 834 834 MANGANESE 3 (BY SIMILARITY).

SQ SEQUENCE 1070 AA; 118441 MW; 4C971B8C42B9C3B5 CRC64;  
 Query Match 1.6%; Score 9; DB 1; Length 1070;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 253 RSLKEEGVQ 261  
 DB 35 RSLKEEGVQ 43  
 RESULT 4  
 CARB METWA  
 ID CARB METWA STANDARD; PRT; 1073 AA.  
 AC P58944;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-  
 DE phosphate synthetase ammonia chain).  
 GN CARB OR MM0038.  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 CC Methanosarcinaceae; Methanosarcina.  
 OK NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goe1 / Goe1 / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RX MEDLINE=22120827; PubMed=12125824;  
 RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,  
 RA Martinez-Arias R., Henne A., Wietzer A., Baumeister S., Jacob C.,  
 RA Bruggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,  
 RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
 RA Fritz H.-J., Gottschalk G.  
 RT "The genome of Methanosarcina mazel: evidence for lateral gene  
 RT transfer between Bacteria and Archaea."  
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +  
 CC phosphate + L-glutamate + carbamoyl phosphate.  
 CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).  
 CC -!- PATHWAY: Arginine biosynthesis.  
 CC -!- SUBUNIT: Pyrimidine biosynthesis; first step.  
 CC promotes the hydrolysis of glutamine to ammonia, which is used by  
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to the carb family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE013223; AM29734.1; --  
 CC HAMAP; MF\_01210; 1;  
 CC InterPro; IPR006275; CarA\_L\_glu.  
 CC InterPro; IPR005483; CPase\_L.  
 CC InterPro; IPR005479; CPase\_L\_D2.  
 CC InterPro; IPR005480; CPase\_L\_D3.  
 CC InterPro; IPR005481; CPase\_L\_N.  
 CC InterPro; IPR004362; MGS like.  
 CC Pfam; PF00289; CPase\_L\_chain; 2.  
 CC Pfam; PF02786; CPase\_L\_D2; 2.  
 CC Pfam; PF02787; CPase\_L\_D3; 1.  
 CC Pfam; PF02142; MGS; 1.  
 CC PRINTS; PR00098; CPASE.  
 CC TIGRFAMs; TIGR01369; CPaseII\_lrg; 1.  
 CC PROSITE; PS00865; CPASE\_1; 1.  
 CC PROSITE; PS00867; CPASE\_2; 1.  
 CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;  
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;  
 FT DOMAIN 1 399  
 FT DOMAIN 400 540 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.  
 FT DOMAIN 541 931 OLIGOMERIZATION DOMAIN.  
 FT DOMAIN 932 1070 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.  
 FT REPEAT 1 540 ALLOSTERIC DOMAIN.  
 FT REPEAT 541 1070  
 FT NP BIND 153 210 ATP (POTENTIAL).  
 FT NP BIND 284 350 ATP (POTENTIAL).  
 FT METAL 284 350 MANGANESE 1 (BY SIMILARITY).  
 FT METAL 296 296 MANGANESE 1 AND 2 (BY SIMILARITY).  
 FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).  
 FT METAL 822 822 MANGANESE 3 (BY SIMILARITY).  
 FT METAL 834 834 MANGANESE 3 (BY SIMILARITY).

KW ATP-binding; Manganese; Complete proteome.  
 FT DOMAIN 1 399 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.  
 FT DOMAIN 400 540 OLIGOMERIZATION DOMAIN.  
 FT DOMAIN 541 931 CARDAMOMYL PHOSPHATE SYNTHETIC DOMAIN.  
 FT DOMAIN 932 1073 ALLOSTERIC DOMAIN.  
 FT REPEAT 1 540  
 FT REPEAT 541 1073  
 FT NP\_BIND 153 210  
 FT NP\_BIND 300 350  
 FT NP\_BIND 284 284  
 FT METAL 296 296  
 FT METAL 298 298  
 FT METAL 822 822  
 FT METAL 834 834  
 SQ SEQUENCE 1073 AA; 118963 MW; ACBD9B5DEFC1EAD1 CRC64;  
 Query Match 1.6%; Score 9; DB 1; Length 1073;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 253 RSLKEGVQ 261  
 DB 35 RSLKEGVQ 43  
 RESULT 5  
 ID IF4G RABIT STANDARD; PRT; 1402 AA.  
 AC P41110;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Eukaryotic translation initiation factor 4 gamma (eIF-4-gamma) (eIF4G)  
 DE (eIF-4G) (p220).  
 GN EIP4G.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 479-500.  
 RC STRAIN=New Zealand white; TISSUE=Brain;  
 RX MEDLINE=93374895; PubMed=8396129;  
 RA Lamphear B.J., Fan X., Yang F., Waters D., Liebig H.-D.,  
 RA Klump H., Kuechler E., Skern T., Rhoads R.E.;  
 RT "Mapping the cleavage site in protein synthesis initiation factor  
 RT eIF-4 gamma of the 2A proteases from human Cocksackievirus and  
 RT rhinovirus."  
 RL J. Biol. Chem. 268:19200-19203(1993).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=93054654; PubMed=1429670;  
 RA Yan R., Rychlik W., Etchison D., Rhoads R.E.;  
 RT "Amino acid sequence of the human protein synthesis initiation factor  
 RT eIF-4 gamma."  
 RL J. Biol. Chem. 267:23226-23231(1992).  
 CC -!- FUNCTION: Component of the protein complex EIF4F, which is  
 CC involved in the recognition of the mRNA cap, ATP-dependent  
 CC unwinding of 5'-terminal secondary structure and recruitment of  
 CC mRNA to the ribosome.  
 CC -!- SUBUNIT: EIF4F is a multi-subunit complex, the composition of  
 CC which varies with external and internal environmental conditions.  
 CC It is composed of at least EIF4A, EIF4E and EIF4G. EIF4G interacts  
 CC with the serine/threonine kinases MNK1 and MNK2. Appears to act  
 CC as a scaffold protein, holding these enzymes in place to  
 CC phosphorylate EIF4E (By similarity).  
 CC -!- SIMILARITY: THE C-TERMINAL REGION IS SIMILAR TO THE N-TERMINAL  
 CC REGION OF WHEAT EUKARYOTIC INITIATION FACTOR (ISO)4F SUBUNIT P62.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; L22090; AAA31242.1; -.  
 DR PIR; I46707; I46707.  
 DR InterPro; IPR008938; ARM  
 DR InterPro; IPR003307; eIF5C.  
 DR InterPro; IPR003890; IF\_eIF4G.  
 DR InterPro; IPR003891; IF\_eIF4G\_MA3.  
 DR Pfam; PF02847; MA3; 1.  
 DR Pfam; PF02854; MIF4G; 1.  
 DR SMART; SM00202; W2; 1.  
 DR SMART; SM00515; eIF5C; 1.  
 DR SMART; SM00544; MA3; 1.  
 DR SMART; SM00543; MIF4G; 1.  
 KW Initiation factor; Protein biosynthesis; Translation regulation;  
 KW Phosphorylation; RNA-binding.  
 FT DOMAIN 188 192 POLY-PRO.  
 FT DOMAIN 262 275 POLY-GLU.  
 FT DOMAIN 1393 1398 POLY-GLU.  
 SQ SEQUENCE 1402 AA; 154050 MW; 7FD85D7E30519230 CRC64;  
 Query Match 1.6%; Score 9; DB 1; Length 1402;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 72 SBELSQLE 80  
 DB 1243 SBELSQLE 1251  
 RESULT 6  
 ID SVE DEIRA STANDARD; PRT; 483 AA.  
 AC Q9RX30;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)  
 DE (GluRS).  
 GN GLTX OR DR0485.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 OX NCBI\_TaxID=1289;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=E1 / ATCC 13939 / DSM 20539 / NCIB 9279;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +  
 CC diphosphate + L-glutamyl-tRNA(Glu).  
 CC -!- SUBUNIT: Monomer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```
CC or send an email to license@sib.ch).
CC -----
CC EMEL; AE001907; AAF10063.1; ALT_INIT.
CC DR PIR; H75514; H75514.
CC DR HSP; P27000; IGLN.
CC DR TIGR; DR0485.
CC DR HAMAP; MP_00022; -.
CC DR InterPro; IPR004527; Glut_bact.
CC DR InterPro; IPR000924; Glu_ERNA-synt_ic.
CC DR InterPro; IPR008925; tRNA-synt_bnd.
CC DR InterPro; IPR001412; tRNA-synt_1.
CC DR Pfam; PF00749; tRNA-synt_ic; 1.
CC DR PRINTS; PR00987; TRNASYNTHGLU.
CC DR TIGRPFAMS; TIGR00464; Glut_bact; 1.
CC DR PROSITE; PS00178; AA TRNA_LIGASE_I; 1.
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC FT SITE 14 24 "HIGH" REGION.
CC FT SITE 253 257 "KMSKS" REGION.
CC FT BINDING 256 256 ATP (BY SIMILARITY).
CC SQ SEQUENCE 483 AA; 54689 MW; 7F6F5C9F3A4567 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 483;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 APPSPRVT 519
DB 3 APPSPRVT 10

RESULT 7
MURD_RICCN STANDARD; PRT; 500 AA.
AC Q92160;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE UPD-N-acetylmuramoylalanine--D-glutamate ligase (EC 6.3.2.9) (UDP-N-
DE acetylmuramoyl-L-alanyl-D-glutamate synthetase) (D-glutamic acid
DE adding enzyme).
GN MURD OR RC0560.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293.2093-2098(2001).
CC -!- FUNCTION: Cell wall formation. Catalyzes the addition of glutamate
CC to the nucleotide precursor UDP-N-acetylmuramoyl-L-alanine (UMA)
CC (by similarity).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanine +
CC glutamate = ADP + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamate.
CC -!- PATHWAY: Peptidoglycan biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- SIMILARITY: Belongs to the murDEP family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
```

```
DR EMEL; AE008617; AAL03098.1; -.
DR PIR; H97769; H97769.
DR HAMAP; MF_00639; atypical; 1.
DR InterPro; IPR000713; Mur_ligase.
DR InterPro; IPR004101; Mur_ligase_C.
DR InterPro; IPR005762; MurD.
DR Pfam; PF01225; Mur_ligase; 1.
DR Pfam; PF02875; Mur_ligase_C; 1.
DR TIGRPFAMS; TIGR01087; murD; 1.
DR ATP-binding; Complete proteome.
KW Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
FT NP_BIND 111 117 ATP (POTENTIAL).
FT DOMAIN 260 306 RPE3.
SQ SEQUENCE 500 AA; 55787 MW; 024B268F6074822C CRC64;

Query Match 1.5%; Score 8; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 QCEPAPPE 108
DB 278 QCEPAPPE 285

RESULT 8
CALD_RAT STANDARD; PRT; 531 AA.
AC Q62736;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Non-muscle caldesmon (CDM) (L-caldesmon).
GN CALDI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., MUTAGENESIS, AND PHOSPHORYLATION SITES.
RC TISSUE=Liver;
RX MEDLINE=95181370; PubMed=7876150;
RA Yamashiro S., Yamakita Y., Yoshida K.-S., Takiguchi K., Matsumura F.;
RT "Characterization of the COOH terminus of non-muscle caldesmon
RT mutants lacking mitosis-specific phosphorylation sites.";
RN J. Biol. Chem. 270:4023-4030(1995).
RL [2]
RP PHOSPHORYLATION BY CDC2.
RX MEDLINE=91095023; PubMed=1986309;
RA Yamashiro S., Yamakita Y., Hosoya H., Matsumura F.;
RT "Phosphorylation of non-muscle caldesmon by p34cdc2 kinase during
RT mitosis.";
RL Nature 349:169-172(1991).
CC -!- FUNCTION: Actin- and myosin-binding protein implicated in the
CC regulation of actomyosin interactions in smooth muscle and actin
CC nonmuscle cells (could act as a bridge between myosin and actin
CC filaments). Stimulates actin binding of tropomyosin which
CC increases the stabilization of actin filament structure. In muscle
CC tissues, inhibits the actomyosin ATPase by binding to P-actin.
CC This inhibition is attenuated by calcium-calmodulin and is
CC potentiated by tropomyosin. Interacts with actin, myosin, two
CC molecules of tropomyosin and with calmodulin. Also play an
CC essential role during cellular mitosis and receptor capping.
CC -!- SUBCELLULAR LOCATION: On thin filaments in smooth muscle and on
CC stress fibers in fibroblasts (nonmuscle) (By similarity).
CC -!- TISSUE SPECIFICITY: High-molecular-weight caldesmon (h-caldesmon)
CC is predominantly expressed in smooth muscles, whereas low-
CC molecular-weight caldesmon (l-caldesmon) is widely distributed in
CC non-muscle tissues and cells. Not expressed in skeletal muscle or
CC heart (By similarity).
CC -!- DOMAIN: The N-terminal part seems to be a myosin/calmodulin-
CC binding domain, and the C-terminal a tropomyosin/actin/calmodulin-
CC binding domain. These two domains are separated by a central
CC helical region in the smooth-muscle form.
```

```

CC CC -!- PTM: In non-muscle cells, phosphorylation by CDC2 during mitosis
CC CC causes caldesmon to dissociate from microfilaments.
CC CC Phosphorylation reduces caldesmon binding to actin, myosin, and
CC CC calmodulin as well as its inhibition of actomyosin ATPase
CC CC activity. Phosphorylation also occurs in both quiescent and
CC CC dividing smooth muscle cells with similar effects on the
CC CC interaction with actin and calmodulin and on microfilaments
CC CC reorganization.
CC CC
CC CC -!- SIMILARITY: Belongs to the caldesmon family.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U18419; AAA68521.1; --
CC CC PIR; A55887; A55887;
CC CC InterPro; IPR006017; Caldesmon_LSP.
CC CC InterPro; IPR006018; Caldesmon_LSP.
CC CC Pfam; PF02029; Caldesmon; 1.
CC CC PRINTS; PR01076; CALDESMON.
CC CC Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
CC CC Alternative splicing.
CC CC
CC CC DOMAIN 20 200 MYOSIN AND CALMODULIN-BINDING (BY
CC CC SIMILARITY).
CC CC FT DOMAIN 303 360
CC CC FT DOMAIN 402 412 TROPOMYOSIN-BINDING (POTENTIAL).
CC CC FT DOMAIN 392 424 TROPOMYOSIN-BINDING (POTENTIAL).
CC CC FT DOMAIN 454 460 STRONG ACTIN-BINDING (BY SIMILARITY).
CC CC FT DOMAIN 506 531 CALMODULIN-BINDING (BY SIMILARITY).
CC CC FT DOMAIN 33 40 WEAK ACTIN-BINDING (BY SIMILARITY).
CC CC FT DOMAIN 180 189 POLY-ARG.
CC CC FT DOMAIN 279 282 POLY-GLU.
CC CC FT DOMAIN 319 322 POLY-ARG.
CC CC FT DOMAIN 336 339 POLY-GLU.
CC CC FT MOD_RES 249 249 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD_RES 462 462 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD_RES 468 468 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD_RES 491 491 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD_RES 497 497 PHOSPHORYLATION (BY CDC2).
CC CC FT MOD_RES 527 527 PHOSPHORYLATION (BY CDC2).
CC CC FT MUTAGEN 249 249 S->A: DECREASES STRONGLY PHOSPHORYLATION-
CC CC FT MUTAGEN 462 462 S->A: DECREASES PHOSPHORYLATION-DEPENDENT
CC CC FT MUTAGEN 468 468 T->A: DECREASES PHOSPHORYLATION-DEPENDENT
CC CC FT MUTAGEN 491 491 ACTIN BINDING.
CC CC FT MUTAGEN 497 497 S->A: DECREASES PHOSPHORYLATION-DEPENDENT
CC CC FT MUTAGEN 497 497 S->A: DECREASES PHOSPHORYLATION-DEPENDENT
CC CC FT MUTAGEN 527 527 S->A: DOES NOT DECREASE PHOSPHORYLATION-
CC CC FT SEQUENCE 531 AA; 60584 MW; CBEC50271A23829 CRC64;
CC CC
CC CC Query Match 1.5%; Score 8; DB 1; Length 531;
CC CC Best Local Similarity 100.0%; Pred.No. 13;
CC CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 263 AREEEKR 270
CC CC DB 334 AREEEKR 341
CC CC
CC CC RESULT 9
CC CC ID_VG50_HSVSA STANDARD; PRT; 535 AA.
CC CC AC Q01012;
CC CC DT 01-APR-1993 (Rel. 25, Created)
CC CC DT 01-APR-1993 (Rel. 25, Last sequence update)

```

```

DT DT 01-APR-1993 (Rel. 25, Last annotation update)
DE DE Probable transcription activator EDRF1.
GN GN 50 OR EDRF1.
OS OS Herpesvirus saimiri (strain 11).
OC OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC OC Gammaherpesvirinae; Rhadinovirus.
OX OX NCBI_TaxID=10383;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RX RX MEDLINE=92333688; PubMed=1321287;
RA RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA RA Honess R.W.;
RT RT "Primary structure of the herpesvirus saimiri genome.";
RL RL J. Virol. 66:5047-5058(1992).
RN RN [2]
RP RP SEQUENCE FROM N.A.
RX RX MEDLINE=92230228; PubMed=1314457;
RA RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT RT organization between HVS and Epstein-Barr virus.";
RL RL Virology 188:296-310(1992).
RN RN [3]
RP RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX RX Nicholas J., Coles L.S., Newman C., Honess R.W.;
RT RT "Regulation of the herpesvirus saimiri (HVS) delayed-early
RT RT 110-kilodalton promoter by HVS immediate-early gene products and a
RT RT homolog of the Epstein-Barr virus R trans activator.";
RL RL J. Virol. 62:2457-2466(1988).
CC CC -!- FUNCTION: TRANSCRIPTION ACTIVATION. REGULATES THE DELAYED-EARLY
CC CC 110 KDA PROMOTER.
CC CC -!- SIMILARITY: TO EBV BRLF1.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; X64346; CAA45672.1; ALT_INIT.
CC CC EMBL; M86409; AAA46124.1; ALT_INIT.
CC CC EMBL; M60850; AAA46159.1; ALT_INIT.
CC CC InterPro; IPR004998; Herpes_TAF50.
CC CC Pfam; PF03326; Herpes_TAF50; 1.
CC CC Transcription regulation; Activator; DNA-binding; Early protein.
CC CC SEQUENCE 535 AA; 60050 MW; B4F2B9ABA38816FB CRC64;
CC CC
CC CC Query Match 1.5%; Score 8; DB 1; Length 535;
CC CC Best Local Similarity 100.0%; Pred.No. 13;
CC CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC CC
CC CC QY 167 GLGKEITL 174
CC CC DB 187 GLGKEITL 194
CC CC
CC CC RESULT 10
CC CC ID_TFC5_YEAST STANDARD; PRT; 594 AA.
CC CC AC P46678;
CC CC DT 01-NOV-1995 (Rel. 32, Created)
CC CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC CC DS Transcription factor TFIIB B* component (TFIIB90).
CC CC TFC5 OR TFC7 OR YNL039W OR N2682.
CC CC Saccharomyces cerevisiae (Baker's yeast).
CC CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CC CC NCBI_TaxID=4932;
CC CC [1]

```

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=9603864; PubMed=7568218;  
RA Kassaravitis G.A., Nguyen S.T., Kobayashi R., Kumar A.,  
RA Glodushchek E.P., Pisano M.;  
RT "Cloning, expression, and function of TFC5, the gene encoding the B'  
RT component of the Saccharomyces cerevisiae RNA polymerase III  
RL transcription factor TFIIB";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:9786-9790(1995).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=96203116; PubMed=8617241;  
RA Rueth J., Conesa C., Dieci G., Lefebvre O., Duesterhoeft A.,  
RA Ottolengo S., Sentenac A.;  
RT "A suppressor of mutations in the class III transcription system  
RT encodes a component of yeast TFIIB";  
RL EMBO J. 15:1941-1949(1996).  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C;  
RX MEDLINE=97313271; PubMed=9169875;  
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgore W.,  
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,  
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,  
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Klein E.,  
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,  
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,  
RA Petal P.X., Pohl T.M., Fumelle D., Schafer M., Scharfe M.,  
RA Schrems B., Schramm S., Schroeder M., Sciu A.M., Tettelin H.,  
RA Urrastarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,  
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,  
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";  
RL Nature 387:103-105(1997).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U31819; AAC49073.1; -  
DR EMBL; U38415; AAC49364.1; -  
DR EMBL; U37533; AAC49348.1; -  
DR EMBL; Z71315; CAA95906.1; -  
DR PIR; S62141; S62141.  
DR GerMOnline; 143046; -  
DR TRAFAC; T00055; -  
DR SGD; S0004984; TFC5.  
DR InterPro; IPR001005; Myb\_DNA\_binding.  
DR Pfam; PF00249; myb\_DNA-binding; 1.  
DR SMART; SM00717; SANT; 1.  
KW Transcription regulation; Activator; Nuclear protein.  
SQ SEQUENCE 594 AA; 67687 MW; 7B0115BBB2491175 CRC64;  
  
Query Match 1.5%; Score 8; DB 1; Length 594;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 265 EEEERKE 272  
DB 334 EEEERKE 341  
|||||  
  
RESULT 11  
YP58 YEAST STANDARD; PRT; 758 AA.  
AC Q92929;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Hypothetical protein YPL158C.  
  
GN OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=97103777; PubMed=8948103;  
RA Furnelle B., Coster F., Goffeau A.;  
RT "The sequence of 55 kb on the left arm of yeast chromosome XVI  
RT identifies a small nuclear RNA, a new putative protein kinase and two  
RL new putative regulators";  
RL Yeast 12:1483-1492(1996).  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=97313271; PubMed=9169875;  
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgore W.,  
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,  
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,  
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Klein E.,  
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,  
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,  
RA Petal P.X., Pohl T.M., Fumelle D., Schafer M., Scharfe M.,  
RA Schrems B., Schramm S., Schroeder M., Sciu A.M., Tettelin H.,  
RA Urrastarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,  
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,  
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";  
RL Nature 387:103-105(1997).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X96770; CAA65563.1; -  
DR EMBL; Z73514; CAA97863.1; -  
DR PIR; S65169; S65169.  
DR GerMOnline; 144140; -  
DR SGD; S0006079; YPL158C.  
KW Hypothetical protein; Coiled coil.  
FT DOMAIN 253 263 POLY-SER.  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 648 711 COILED COIL (POTENTIAL).  
FT DOMAIN 652 658 POLY-GLU.  
FT DOMAIN 664 669 POLY-GLU.  
FT DOMAIN 686 690 POLY-GLU.  
SQ SEQUENCE 758 AA; 84845 MW; 3B4FA92B91C87F2B CRC64;  
  
Query Match 1.5%; Score 8; DB 1; Length 758;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 265 EEEERKE 272  
DB 687 EEEERKE 694  
|||||  
  
RESULT 12  
RPOB LOTJA STANDARD; PRT; 1070 AA.  
ID RPOB LOTJA  
AC Q9B559;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)



10-OCT-2003 (Rel. 42, Last annotation update)  
DNA-directed RNA polymerase beta chain (EC 2.7.7.6).  
RPOB  
Lotus japonicus.  
Chloroplast.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
eurosids I; Fabiales; Fabaceae; Papilionoideae; Lotaeae; Lotus.  
NCBI\_TaxID=34305;  
(1)  
SEQUENCE FROM N.A.  
STRAIN=Accession MG-20;  
MEDLINE=21082929; PubMed=11214967;  
Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;  
"Complete structure of the chloroplast genome of a legume, Lotus  
japonicus";  
DNA Res. 7:323-330(2000).  
-!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription  
of DNA into RNA using the four ribonucleoside triphosphates as  
substrates.  
-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
{RNA}(N).  
-!- SUBUNIT: In chloroplasts, the RNA polymerase is composed of four  
subunits: alpha, beta, beta', and beta".  
-!- SUBCELLULAR LOCATION: Chloroplast.  
-!- SIMILARITY: Belongs to the RNA polymerase beta chain family.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF002983; BAB3194.1; -.  
HSP; O9KMU7; 1HOW.  
InterPro: IPR007121; RNA pol B.  
InterPro: IPR007644; RNA pol Rpb2 1.  
InterPro: IPR007642; RNA pol Rpb2 2.  
InterPro: IPR007645; RNA pol Rpb2 3.  
InterPro: IPR007120; RNA pol Rpb2 6.  
InterPro: IPR007641; RNA pol Rpb2 7.  
InterPro: IPR007641; RNA pol Rpb2 7.  
Pfam: PF04563; RNA pol Rpb2 1; 1.  
Pfam: PF04561; RNA pol Rpb2 2; 1.  
Pfam: PF04565; RNA pol Rpb2 3; 1.  
Pfam: PF00562; RNA pol Rpb2 6; 1.  
Pfam: PF04560; RNA pol Rpb2 7; 1.  
PROSITE; PS01166; RNA POL BETA; 1.  
TRANSFERASE; DNA-directed RNA polymerase; Transcription; Chloroplast.  
SEQUENCE 1070 AA; 120892 MW; 7126BD99B0B5C063 CRC64;  
Query Match 1.5%; Score 8; DB 1; Length 1070;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 242 ESLCRELQ 249  
Db 241 ESLCRELQ 248  
RESULT 13  
CARB\_THETN STANDARD; PRT; 1072 AA.  
AC Q8R8K0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-  
phosphate synthetase ammonia chain).  
GN CARB OR TTR0816.  
OS Thermoanaerobacter tengcongensis.  
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;

Thermoanaerobacteriaceae; Thermoanaerobacter.  
NCBI\_TaxID=119072;  
(1)  
SEQUENCE FROM N.A.  
STRAIN=M54 / JCM 11007;  
MEDLINE=2192816; PubMed=11997336;  
Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
Tan H., Chen R., Wang J., Yu J., Yang H.;  
"A complete sequence of T. tengcongensis genome";  
Genome Res. 12:689-700(2002).  
-!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +  
phosphate + L-glutamate + carbamoyl phosphate.  
-!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).  
-!- PATHWAY: Arginine biosynthesis.  
-!- SUBUNIT: Pyrimidine biosynthesis; first step.  
-!- FUNCTION: Composed of two chains; the small (or glutamine) chain  
promotes the hydrolysis of glutamine to ammonia, which is used by  
the large (or ammonia) chain to synthesize carbamoyl phosphate (By  
similarity).  
-!- SIMILARITY: Belongs to the carb family.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AE013048; AAM24073.1; -.  
HAMAP; MP 01210; -; 1.  
InterPro: IPR006275; CarA\_L\_glu.  
InterPro: IPR005483; CPase\_L\_D2.  
InterPro: IPR005479; CPase\_L\_D3.  
InterPro: IPR005480; CPase\_L\_D3.  
InterPro: IPR005481; CPase\_L\_N.  
InterPro: IPR004362; MGS like.  
Pfam: PF00289; CPase\_L\_D2; 2.  
Pfam: PF02786; CPase\_L\_D2; 2.  
Pfam: PF02787; CPase\_L\_D3; 1.  
Pfam: PF02142; MGS; 1.  
PRINTS; PR00098; CPASE.  
TIGR; TIGR01369; CPase1\_lrg; 1.  
PROSITE; PS00866; CPASE 1; 1.  
PROSITE; PS00867; CPASE 2; 2.  
Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;  
ATP-binding; Manganese; Complete proteome.  
DOMAIN 1 401  
DOMAIN 402 546  
DOMAIN 547 929  
DOMAIN 930 1072  
CARBOXYPHOSPHATE SYNTHETIC DOMAIN.  
OLIGOMERIZATION DOMAIN.  
CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.  
ALLOSTERIC DOMAIN.  
ATP (POTENTIAL).  
ATP (POTENTIAL).  
MANGANESE 1 (BY SIMILARITY).  
MANGANESE 2 (BY SIMILARITY).  
MANGANESE 3 (BY SIMILARITY).  
MANGANESE 3 (BY SIMILARITY).  
MANGANESE 3 (BY SIMILARITY).  
SEQUENCE 1072 AA; 119000 MW; D5F08ADC8BE75D37 CRC64;  
Query Match 1.5%; Score 8; DB 1; Length 1072;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 254 SLKEEGVQ 261  
Db 36 SLKEEGVQ 43  
RESULT 14

[illegible]

FT DOMAIN	742	763	IQ 1.
FT DOMAIN	764	787	IQ 2.
FT DOMAIN	788	817	IQ 3.
FT DOMAIN	1206	1304	PH 1.
FT DOMAIN	1386	1491	PH 2.
FT DOMAIN	1694	2038	FERM.
FT DOMAIN	800	941	COILED COIL (POTENTIAL).
SQ SEQUENCE	2052 AA;	235837 MW;	43DF13424B4B2D28 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 2052;  
Best Local Similarity 100.0%; Pred.No. 41;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	264	REBEKRX	271
Db	824	REBEKRX	831

Search completed: June 8, 2004, 16:38:39  
Job time : 35 secs



is claimed (see AAB82820). Methods of determining if an animal is at risk for atherosclerosis, methods for evaluating an agent for use in treating atherosclerosis, and methods for treating a cell having an abnormality in structure or metabolism of LBP are claimed. Pharmaceutical compositions comprising an LBP polypeptide or nucleic acid, and vaccine compositions, are also claimed

Sequence 546 AA;

Query Match 100.0%; Score 2785; DB 4; Length 546;  
Best Local Similarity 100.0%; Pred. No. 5.7e-160;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKNQDKKNGAKQSNPKSSPCQAGPAGQERPSQAPAVEAGPSSQAPRKPGEQA 60  
Db 1 MKNQDKKNGAKQSNPKSSPCQAGPAGQERPSQAPAVEAGPSSQAPRKPGEQA 60  
Qy 61 RTAQSGALRDVSEBLSRQLEDILSTYCVNNQGGPGDGAQEPAPEDAEKSTTYVARN 120  
Db 61 RTAQSGALRDVSEBLSRQLEDILSTYCVNNQGGPGDGAQEPAPEDAEKSTTYVARN 120  
Qy 121 GEPEPTPVNGEKPSKGDPTNTEIROSDEVDGDRHRRPQKKAAGLGEKITILLMOTLN 180  
Db 121 GEPEPTPVNGEKPSKGDPTNTEIROSDEVDGDRHRRPQKKAAGLGEKITILLMOTLN 180  
Qy 181 TLSTPEEKLAALCKKYAELLEHRNSQKMLLOKQSLVOEKDHLRGHSHKAVLARSK 240  
Db 181 TLSTPEEKLAALCKKYAELLEHRNSQKMLLOKQSLVOEKDHLRGHSHKAVLARSK 240  
Qy 241 LESLCRELQNRNLSKEGVORAREBEKREKVTSHFQVTLNDIQLQMEQHNERNKLRQ 300  
Db 241 LESLCRELQNRNLSKEGVORAREBEKREKVTSHFQVTLNDIQLQMEQHNERNKLRQ 300  
Qy 301 ENMELAREKKLISQYELREHIDKVPKHDLQQLVDKLLQQAQEMLKAEERHOREKD 360  
Db 301 ENMELAREKKLISQYELREHIDKVPKHDLQQLVDKLLQQAQEMLKAEERHOREKD 360  
Qy 361 PLLKEAVESQRMCELMQQTTHLKOQALYTEKPEEPQNTLSKSSVFYTFKQEMEKWTK 420  
Db 361 PLLKEAVESQRMCELMQQTTHLKOQALYTEKPEEPQNTLSKSSVFYTFKQEMEKWTK 420  
Qy 421 KIKLEKETMYRWRSSNKLALMAEKTVDKLEGLQVKIQRLEKLCRALQTERND 480  
Db 421 KIKLEKETMYRWRSSNKLALMAEKTVDKLEGLQVKIQRLEKLCRALQTERND 480  
Qy 481 LNKRVQDLSAGCGSLTDSGPERPFGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQE 540  
Db 481 LNKRVQDLSAGCGSLTDSGPERPFGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQE 540  
Qy 541 PTSARA 546  
Db 541 PTSARA 546

RESULT 2

ID ABR41210 standard; protein; 639 AA.

XX ABR41210;

AC ABR41210;

XX 02-JUN-2003 (first entry)

XX Human DITHP extracellular signalling protein.

XX Human, dithp; diagnostic and therapeutic polynucleotide; diagnosis;

KW cancer; cell proliferative disorder; autoimmune disorder;

KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;

KW neurological disorder; gastrointestinal disorder; transport disorder;

KW connective tissue disorder; drug screening; proteome analysis;

KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;

KW disease model; toxicological testing; transcript imaging;

KW extracellular signalling.

OS Homo sapiens.

PN WO200297031-A2.

XX 05-DEC-2002.

XX 27-MAR-2002; 2002WO-US010056.

XX 28-MAR-2001; 2001US-0279619P.

XX 29-MAR-2001; 2001US-0280067P.

XX 29-MAR-2001; 2001US-0280068P.

XX 16-MAY-2001; 2001US-0291280P.

XX 17-MAY-2001; 2001US-0291829P.

XX 17-MAY-2001; 2001US-0291849P.

XX 19-JUN-2001; 2001US-0299428P.

XX 20-JUN-2001; 2001US-0299776P.

XX 20-JUN-2001; 2001US-0300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

XX Dufour GB, Hillman JL, Yu JY, Tuason O, Yap PE, Amshy SR;

XX Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX WPI; 2003-129518/12.

XX N-PSDB; ACC66153.

XX Novel human diagnostic and therapeutic polypeptide useful for identifying

XX test compound which specifically binds to a polypeptide encoded by human

XX diagnostic and therapeutic polynucleotide, and to induce antibodies.

XX Claim 27; SEQ ID NO 745; 591pp; English.

XX The invention relates to novel human diagnostic and therapeutic

XX polynucleotides designated dithp (ACC46080-ACC46749), and to their encoded

XX proteins (DITHP; ABR41136-ABR41182). The invention also relates to

XX polynucleotide sequences at least 90% identical to the dithp cDNA

XX sequences of the invention; recombinant vectors, host cells and

XX transgenic organisms comprising a dithp nucleic acid sequence; the

XX recombinant production of DITHP proteins; antibodies specific for DITHP

XX proteins; microarrays comprising dithp nucleic acid sequences; methods of

XX detecting dithp nucleotide and protein sequences; methods of screening

XX for compounds which specifically bind a DITHP protein; and methods of

XX assessing the toxicity of test compounds using a dithp hybridisation

XX probe. Dithp nucleic acid sequences and DITHP proteins may be used in the

XX diagnosis of a wide variety of conditions including cancer and other cell

XX proliferative disorders; autoimmune or inflammatory disorders; bacterial,

XX viral, fungal or parasitic infections; hormonal disorders; metabolic

XX disorders; neurological disorders; gastrointestinal disorders; transport

XX disorders; and connective tissue disorders. They may also be used to

XX screen for modulators of protein activity or gene expression. DITHP

XX proteins can additionally be used in analysis of the proteome of a tissue

XX or cell type and to induce antibodies. The dithp nucleic acids are

XX additionally useful in somatic or germline gene therapy of the disorders

XX mentioned above, as a source of antisense sequences, as a source of

XX probes and primers, in genotyping and identification of individuals, in

XX the generation of transgenic animal models of human disease or knock in

XX humanised animals, in toxicological testing, and in transcript imaging.

XX The present sequence represents a DITHP protein which has extracellular

XX signalling activity. Note: The sequence data for this patent did not form

XX part of the printed specification, but was obtained in electronic format

XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 639 AA;

XX Query Match 100.0%; Score 2785; DB 6; Length 639;

XX Best Local Similarity 100.0%; Pred. No. 6.8e-160;

XX Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 MKNQDKKNGAKQSNPKSSPCQAGPAGQERPSQAPAVEAGPSSQAPRKPGEQA 60

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 94 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEQA 153  
Qy 61 RTAQSALRDVSEELSRQLEDILSTYVDNNQGGGEGDGAQGEPAEPDAEKSRTYVARN 120  
Db 154 RTAQSALRDVSEELSRQLEDILSTYVDNNQGGGEGDGAQGEPAEPDAEKSRTYVARN 213  
Qy 121 GPEPTPVVNGEKEPSKSGDPNTEIEIRQSDVEGDRDHRPQEKKAAGLKGKTTLLMQTLN 180  
Db 214 GPEPTPVVNGEKEPSKSGDPNTEIEIRQSDVEGDRDHRPQEKKAAGLKGKTTLLMQTLN 273  
Qy 181 TLSTPEKLAALCKKYAELLEBEHNSQKMKLQKQSOLVQEKDHLRGHSKAVLARSK 240  
Db 274 TLSTPEKLAALCKKYAELLEBEHNSQKMKLQKQSOLVQEKDHLRGHSKAVLARSK 333  
Qy 241 LESLCELOHNRSLKEGVQVAREBEERKEVTSFQVTLNDIOLQMEQHNRNSKLQ 300  
Db 334 LESLCELOHNRSLKEGVQVAREBEERKEVTSFQVTLNDIOLQMEQHNRNSKLQ 393  
Qy 301 ENWELAEALRLKQIEQVELBEHIDKVFVKHDIQQQLVDKLAQOAOBMLKEAERHQREK 360  
Db 394 ENWELAEALRLKQIEQVELBEHIDKVFVKHDIQQQLVDKLAQOAOBMLKEAERHQREK 453  
Qy 361 FLKKEAVESORWELMKQOETHLKOALALYTKFEFEFONTLSSSEVFTTFKQEMERMTK 420  
Db 454 FLKKEAVESORWELMKQOETHLKOALALYTKFEFEFONTLSSSEVFTTFKQEMERMTK 513  
Qy 421 KIKLEKETTMVRSRWESSNKALLENMAEKTVRDKELEGLOVKIQRLKLCRALQTERND 480  
Db 514 KIKLEKETTMVRSRWESSNKALLENMAEKTVRDKELEGLOVKIQRLKLCRALQTERND 573  
Qy 481 LNKRVODLSAGGQSLTDSGPRRPGCAOAPSSPRVTEAPCYGAPSTEASGTGPGQE 540  
Db 574 LNKRVODLSAGGQSLTDSGPRRPGCAOAPSSPRVTEAPCYGAPSTEASGTGPGQE 633  
Qy 541 PPSARA 546  
Db 634 PPSARA 639

RESULT 3  
ABR41186  
ID ABR41186 standard; protein; 639 AA.  
XX ABR41186;  
AC ABR41186;  
XX ABR41186;  
DT 02-JUN-2003 (first entry)  
DE Human DITHP extracellular signalling protein.  
XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;  
KW cancer; cell proliferative disorder; autoimmune disorder;  
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;  
KW neurological disorder; gastrointestinal disorder; transport disorder;  
KW connective tissue disorder; drug screening; proteome analysis;  
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;  
KW disease model; toxicological testing; transcript imaging;  
KW extracellular signalling.  
XX Homo sapiens.  
OS  
XX WO200297031-A2.  
PN  
XX  
XX 05-DEC-2002.  
XX  
XX 27-MAR-2002; 2002WO-US010056.  
XX  
XX 28-MAR-2001; 2001US-0279619P.  
PR 29-MAR-2001; 2001US-0280067P.  
PR 29-MAR-2001; 2001US-0280068P.  
PR 16-MAY-2001; 2001US-0291280P.  
PR 17-MAY-2001; 2001US-0291829P.  
PR 17-MAY-2001; 2001US-0291849P.  
PR 19-JUN-2001; 2001US-0299428P.

PR 20-JUN-2001; 2001US-0299776P.  
PR 20-JUN-2001; 2001US-030001P.  
PA (INCY-) INCYTE GENOMICS INC.  
PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;  
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
PI Flores V, Marvaha R, Lo A, Lan RY, Urashka ME;  
XX WPI; 2003-129518/12.  
DR N-PSDB; ACC46130.  
XX  
PT Novel human diagnostic and therapeutic polypeptide useful for identifying  
PT test compound which specifically binds to a polypeptide encoded by human  
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.  
XX  
PS Claim 27; SEQ ID NO 721; 591pp; English.  
CC The invention relates to novel human diagnostic and therapeutic  
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded  
CC proteins (DITHP: ABR41136-ABR41812). The invention also relates to  
CC polynucleotide sequences at least 90% identical to the dithp cDNA  
CC sequences of the invention; recombinant vectors, host cells and  
CC transgenic organisms comprising a dithp nucleic acid sequence; the  
CC recombinant production of DITHP proteins; antibodies specific for DITHP  
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of  
CC detecting dithp nucleotide and protein sequences; methods of screening  
CC for compounds which specifically bind a DITHP protein; and methods of  
CC assessing the toxicity of test compounds using a dithp hybridisation  
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the  
CC diagnosis of a wide variety of conditions including cancer and other cell  
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,  
CC viral, fungal or parasitic infections; gastrointestinal disorders; transport  
CC disorders; neurological disorders; hormonal disorders; metabolic  
CC disorders; and connective tissue disorders. They may also be used to  
CC screen for modulators of protein activity or gene expression. DITHP  
CC proteins can additionally be used in analysis of the proteome of a tissue  
CC or cell type and to induce antibodies. The dithp nucleic acids are  
CC additionally useful in somatic or germline gene therapy of the disorders  
CC mentioned above, as a source of antisense sequences, as a source of  
CC probes and primers, in genotyping and identification of individuals, in  
CC the generation of transgenic animal models of human disease or knock in  
CC humanised animals, in toxicological testing, and in transcript imaging.  
CC The present sequence represents a DITHP protein which has extracellular  
CC signalling activity. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 639 AA;  
Query Match 100.0%; Score 2785; DB 6; Length 639;  
Best Local Similarity 100.0%; Pred. No. 6.8e-160; Indels 0; Gaps 0;  
Matches 546; Conservative 0; Mismatches 0;  
Qy 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEQA 60  
Db 94 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPRKPGEQA 153  
Qy 61 RTAQSALRDVSEELSRQLEDILSTYVDNNQGGGEGDGAQGEPAEPDAEKSRTYVARN 120  
Db 154 RTAQSALRDVSEELSRQLEDILSTYVDNNQGGGEGDGAQGEPAEPDAEKSRTYVARN 213  
Qy 121 GPEPTPVVNGEKEPSKSGDPNTEIEIRQSDVEGDRDHRPQEKKAAGLKGKTTLLMQTLN 180  
Db 214 GPEPTPVVNGEKEPSKSGDPNTEIEIRQSDVEGDRDHRPQEKKAAGLKGKTTLLMQTLN 273  
Qy 181 TLSTPEKLAALCKKYAELLEBEHNSQKMKLQKQSOLVQEKDHLRGHSKAVLARSK 240  
Db 274 TLSTPEKLAALCKKYAELLEBEHNSQKMKLQKQSOLVQEKDHLRGHSKAVLARSK 333  
Qy 241 LESLCELOHNRSLKEGVQVAREBEERKEVTSFQVTLNDIOLQMEQHNRNSKLQ 300



PR	14-JUL-2000; 2000US-00616289.	AAW49039 standard; protein; 557 AA.
XX	(BOST-) BOSTON HEART FOUND INC.	AAW49039;
PA	Lees AM, Lees RS, Law SW, Arjona AA;	09-NOV-1998 (first entry)
PI	WPI; 2001-565505/63.	Rabbit low density lipoprotein binding protein LBP-3.
DR	N-PSDB; AAH26496.	Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;
XX	New isolated low density lipoprotein binding polypeptide for treating,	receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.
PT	diagnosing and/or identifying therapeutic agents for atherosclerosis.	
XX	Claim 13(h); Fig 8B; 143pp; English.	Oryctolagus cuniculus.
PS		
XX	The present sequence is that of the N-terminal portion of novel human low	
CC	density lipoprotein binding protein 3 (LBP-3). The amino acid sequence is	
CC	deduced from an isolated partial cDNA clone (see AAH26494). A full-length	
CC	sequence is given in AAB2808. Human LBP-3 is an example of claimed LBP	
CC	polypeptides of the invention that are capable of binding to native and	
CC	methylated low density lipoproteins. Also claimed are biologically active	
CC	fragments and analogues of LBPs, polynucleotides encoding LBPs, as well	
CC	as expression vectors, cells and methods of producing the LBPs. Methods	
CC	of determining if an animal is at risk for atherosclerosis, methods for	
CC	evaluating an agent for use in treating atherosclerosis, and methods for	
CC	treating a cell having an abnormality in structure or metabolism of LBP	
CC	are claimed. Pharmaceutical compositions comprising an LBP polypeptide or	
CC	nucleic acid, and vaccine compositions, are also claimed	
XX		
SQ	Sequence 530 AA;	
Query Match 96.7%; Score 2693; DB 4; Length 530;		
Best Local Similarity 99.8%; Pred. No. 2e-154; Indels 0; Gaps 0;		
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	17 KSSPGQPEAGPEGAQERPSQAAPAVAEAGSGSQAPRKPEGAQARTAQSGALRDVSEELS 76	
DB	1 KSSPGQPEAGPEGAQERPSQAAPAVAEAGSGSQAPRKPEGAQARTAQSGALRDVSEELS 60	
QY	77 RQLEDILSTYCVDNNGCGPEDGAGPEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 136	
DB	61 RQLEDILSTYCVDNNGCGPEDGAGPEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAEPAE 120	
QY	137 KGPNTPEIRQSDVEGDRHRRPQEKKAAGLCKEITLLMOTLNTLSTPEKLAALCKKY 196	
DB	121 KGPNTPEIRQSDVEGDRHRRPQEKKAAGLCKEITLLMOTLNTLSTPEKLAALCKKY 180	
QY	197 AELLEHRNSQKQKLLQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 256	
DB	181 AELLEHRNSQKQKLLQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 240	
QY	257 EGVQVAREEERKKEVTSHFVTLNDIQLQMEQHNERNKSLQENNELAERLKLIEQY 316	
DB	241 EGVQVAREEERKKEVTSHFVTLNDIQLQMEQHNERNKSLQENNELAERLKLIEQY 300	
QY	317 ELREEHIDKVRKHDLOQQLVDKLAQQAQEMLEAEERHQEKDFLLKEAVESQRMCELM 376	
DB	301 ELREEHIDKVRKHDLOQQLVDKLAQQAQEMLEAEERHQEKDFLLKEAVESQRMCELM 360	
QY	377 KQETHLKOQALYTEREPFQNTLSKSVFTTFKQEMKTKKIKLKEKETTMYRSRW 436	
DB	361 KQETHLKOQALYTEREPFQNTLSKSVFTTFKQEMKTKKIKLKEKETTMYRSRW 420	
QY	437 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLGAGGQSL 496	
DB	421 ESSNKALLEMAEKTVDKLEGLQVKIQRLKLCRALQTERNDLNKRVQDLGAGGQSL 480	
QY	497 TDGSPERPPEGAQAPSSPRVTAPCYPCGAPSTASGQGPQEPPTSARA 546	
DB	481 TDGSPERPPEGAQAPSSPRVTAPCYPCGAPSTASGQGPQEPPTSARA 530	
RESULT 6		
AAW49039		

ID	AAW49039 standard; protein; 557 AA.	
XX	AAW49039;	
AC	09-NOV-1998 (first entry)	
XX	Rabbit low density lipoprotein binding protein LBP-3.	
DE	Low density lipoprotein binding protein; LDL binding protein 3; LBP-3;	
XX	receptor; rabbit; atherosclerosis; diagnosis; therapy; vaccine.	
KW	Oryctolagus cuniculus.	
XX		
OS		
XX	Key	Location/Qualifiers
PH	Peptide	96.110
FT		/note= "Claim 2"
XX	WO9823282-A1.	
PN	04-JUN-1998.	
XX	26-NOV-1997; 97WO-US021857.	
PD	27-NOV-1996; 96US-0031930P.	
XX	03-JUN-1997; 97US-0048547P.	
PF	(BOST-) BOSTON HEART FOUND INC.	
PR	Lees AM, Lees RS, Law SW, Arjona AA;	
BR	WPI; 1998-322455/28.	
XX	N-PSDB; AAV32836.	
DR	Nucleic acid encoding low density lipoprotein binding proteins and	
XX	related vectors - transformed cells, proteins, and modulators of binding,	
PT	useful for treatment and diagnosis of atherosclerosis and for identifying	
PT	subjects at risk.	
XX	Claim 1; Fig 5; 47pp; English.	
PS	This polypeptide comprises novel rabbit low density lipoprotein (LDL)	
CC	binding protein LBP-3 that is capable of binding both native and methyl	
CC	LDL. Its amino acid sequence was deduced from rabbit abdominal aorta cDNA	
CC	(see AAV32836). cDNA clones (see AAV32834-39) and encoded rabbit and human	
CC	LBPs (see AAW49037-42) are claimed. An abnormality in an aspect of LBP	
CC	metabolism or structure is diagnostic of a risk for atherosclerosis. The	
CC	invention provides: methods for determining if an animal is at risk for	
CC	atherosclerosis (e.g. for prenatal screening); methods for treating	
CC	atherosclerosis (including gene therapy) using e.g. LBP polypeptides to	
CC	bind LDL and thereby prevent formation of atherosclerotic plaque; and	
CC	methods for treating a cell having an abnormality in LBP structure or	
CC	metabolism. Pharmaceutical and vaccine compositions are also provided, as	
CC	well as recombinant vectors and host cells used to produce recombinant	
CC	LBP	
XX		
SQ	Sequence 557 AA;	
Query Match 91.7%; Score 2552.5; DB 2; Length 557;		
Best Local Similarity 91.0%; Pred. No. 6.2e-146;		
Matches 508; Conservative 12; Mismatches 25; Indels 13; Gaps 3;		
QY	1 MKNQDKKNGAAQKSPKSPGQPEAGPEGAQERPSQAAPAVAEAGSGSQAPRKPEGAQA 60	
DB	1 MKNQDKKNGAAQKSPKSPGQPEAGPEGAQERPSQAAPAVAEAGSGSQAPRKPEGAQA 59	
QY	61 RTAQSGALRDVSEELSRLQLEDILSTYCVDNNGCGPEDGAGPEPAEPAEPAEPAEPAE 120	
DB	60 RTAQSGALRDVSEELSRLQLEDILSTYCVDNNGCGPEDGAGPEPAEPAEPAEPAEPAE 119	
QY	121 GSEPEP-TPVWNGEKPKSGDPNTEIRQSDVEGDRHRRPQEKKAAGLCKEITLLMOTL 179	
DB	120 GSEPEP-TPVWNGEKPKSGDPNTEIRQSDVEGDRHRRPQEKKAAGLCKEITLLMOTL 179	



```

QY 180 NTLSTPEKLAALCKKYAEILBEHRNSQKMKLLQKQSLVQKDHILRGHSHKAVLARS 239
DB 180 NTLSTPEKLAALCKKYAEILBEHRNSQKMKLLQKQSLVQKDHILRGHSHKAVLARS 239
QY 240 KLSLCRLQHRNSLKEEGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLR 299
DB 240 KLSLCRLQHRNSLKEEGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLR 299
QY 300 QENNELAERLKLIEQYELREHIDKVFKHDLQQLVDAKLQQAQEMLKEAERHOREK 359
DB 300 QENNELAERLKLIEQYELREHIDKVFKHDLQQLVDAKLQQAQEMLKEAERHOREK 359
QY 360 DFLKKEAVESQRMCELMKQOETHLQKQALYTERKFEFQNTLSKSSSEVFTTFKQEMKMT 419
DB 360 DFLKKEAVESQRMCELMKQOETHLQKQALYTERKFEFQNTLSKSSSEVFTTFKQEMKMT 419
QY 420 KKI KLEKETTYRSRWESSNKALLEMAEBEKTVDKLEGLQVKIQRLEKLCRALQTERN 479
DB 420 KKI KLEKETTYRSRWESSNKALLEMAEBEKTVDKLEGLQVKIQRLEKLCRALQTERN 479
QY 480 DLNKRVDLSAGGQSLTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGAP 528
DB 480 DLNKRVDLSAGGQSLTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGAP 528
QY 529 STEASGQTGPQEPPTATTA 546
DB 540 STEASGQTGPQEPPTATTA 557

```

RESULT 7  
ID AAB82801 standard; protein; 557 AA.

AC AAB82801;  
XX  
XX 12-NOV-2001 (first entry)  
XX Rabbit low density lipoprotein binding protein 3 (LBP-3).  
XX  
XX Low density lipoprotein binding protein 3; LBP-3; LDL; rabbit;  
KW atherosclerosis; antiatherosclerotic; therapy; diagnosis; vaccine.  
XX  
XX Oryctolagus cuniculus.  
XX  
XX WO200164874-A2.

XX  
XX 07-SEP-2001.  
XX 28-FEB-2001; 2001WO-US006356.  
XX  
XX 02-MAR-2000; 2000US-00517849.  
XX 14-JUL-2000; 2000US-00616289.  
XX  
XX (BOST-) BOSTON HEART FOUND INC.

XX Lees AM, Lees RS, Law SW, Arjona AA;  
XX  
XX WPI; 2001-565505/53.  
XX N-PSDB; AAB26491.  
XX  
XX New isolated low density lipoprotein binding polypeptide for treating,  
XX diagnosing and/or identifying therapeutic agents for atherosclerosis.  
XX  
XX Claim 13(e); Fig 5; 143pp; English.

XX The present sequence is that of novel rabbit low density lipoprotein  
XX binding protein 1 (LBP-3). The amino acid sequence is deduced from an  
XX isolated cDNA clone (see AAB26491). Rabbit LBP-3 is an example of claimed  
XX polypeptides of the invention, termed LBPs, that are capable of binding  
XX to native and methylated low density lipoproteins. Also claimed are  
XX biologically active fragments and analogues of LBPs, polynucleotides  
XX encoding LBPs, as well as expression vectors, cells and methods of  
XX producing the LBPs. A polypeptide having the amino acid residues 96-110

CC of the present sequence is claimed (see AAB82819). Methods of determining  
CC if an animal is at risk for atherosclerosis, methods for evaluating an  
CC agent for use in treating atherosclerosis, and methods for treating a  
CC cell having an abnormality in structure or metabolism of LBP are claimed.  
CC Pharmaceutical compositions comprising an LBP polypeptide or nucleic  
CC acid, and vaccine compositions, are also claimed  
XX  
SQ Sequence 557 AA;

```

Query Match 91.7%; Score 2552.5; DB 4; Length 557;
Best local Similarity 91.0%; Pred. NO. 6.2e-146; Indels 13; Gaps 3;
Matches 508; Conservative 12; Mismatches 25;
QY 1 MNQDKKNGAAKQNPKNKSPGQPEAGPEGAQEPSPQAAPAVEAGPGSSQAPRKPEGAQA 60
DB 1 MNQDKKNGAAKQNPKNKSPGQPEAGPEGAQEPSPQAAPAVEAGPGSSQAPRKPEGAQA 59
QY 61 RTQSGALRDVSELGRQLEDILSTVCVNNQGGPCEGAGCEPAPDAKSKRTYVARN 120
DB 60 KTAQPGALCDVSELGRQLEDILSTVCVNNQGGPCEGAGCEPAPDAKSKRTYVARN 119
QY 121 GEPEP-TPVVGNGEKPSPKDPNTEIRQSDVEGDRHRRPQEKKAKGLGKEITLLMQTL 179
DB 120 GEPEP-TPVVGNGEKPSPKDPNTEIRQSDVEGDRHRRPQEKKAKGLGKEITLLMQTL 179
QY 180 NTLSTPEKLAALCKKYAEILBEHRNSQKMKLLQKQSLVQKDHILRGHSHKAVLARS 239
DB 180 NTLSTPEKLAALCKKYAEILBEHRNSQKMKLLQKQSLVQKDHILRGHSHKAVLARS 239
QY 240 KLSLCRLQHRNSLKEEGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLR 299
DB 240 KLSLCRLQHRNSLKEEGVQVAREEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLR 299
QY 300 QENNELAERLKLIEQYELREHIDKVFKHDLQQLVDAKLQQAQEMLKEAERHOREK 359
DB 300 QENNELAERLKLIEQYELREHIDKVFKHDLQQLVDAKLQQAQEMLKEAERHOREK 359
QY 360 DFLKKEAVESQRMCELMKQOETHLQKQALYTERKFEFQNTLSKSSSEVFTTFKQEMKMT 419
DB 360 DFLKKEAVESQRMCELMKQOETHLQKQALYTERKFEFQNTLSKSSSEVFTTFKQEMKMT 419
QY 420 KKI KLEKETTYRSRWESSNKALLEMAEBEKTVDKLEGLQVKIQRLEKLCRALQTERN 479
DB 420 KKI KLEKETTYRSRWESSNKALLEMAEBEKTVDKLEGLQVKIQRLEKLCRALQTERN 479
QY 480 DLNKRVDLSAGGQSLTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGAP 528
DB 480 DLNKRVDLSAGGQSLTDSGPERRP-----EGPGAQAPSSPRVTEAPCYGAP 528
QY 529 STEASGQTGPQEPPTATTA 546
DB 540 STEASGQTGPQEPPTATTA 557

```

RESULT 8  
ABG00839  
ID ABG00839 standard; protein; 1749 AA.  
XX  
XX AC ABG00839;  
XX  
XX 13-FEB-2002 (first entry)  
XX Novel human diagnostic protein #830.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX

PF 30-MAR-2001; 2001WO-US028631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSB-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS65026.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 31198; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_ppt\_sequences  
XX  
SQ Sequence 1749 AA;  
  
Query Match 87.3%; Score 2430.5; DB 4; Length 1749;  
Best Local Similarity 84.9%; Pred. No. 5.3e-138;  
Matches 500; Conservative 4; Mismatches 30; Indels 55; Gaps 6;  
  
Qy 1 MNQDKKNGAAKQSNPKSSPCQPEAGPEGAGQERPSQAAPAVEAGPGSSQAPRKPEGAQA 60  
Db 152 MNQDKKNGAAKQSNPKSSPCQPEAGPEGAGQERPSQAAPAVEAGPGSSQAPRKPEGAQA 211  
  
Qy 61 RTAQSGLRDVSELSRQLEDILSTYCVNNQGGGEGDGAQGPAPEDAEKSTTYARN 120  
Db 212 RTAQSGLRDVSELSRQLEDILSTYCVNNQGGGEGDGAQGPAPEDAEKSTTYARN 271  
  
Qy 121 GEPEPTFWNGEKSPKSGDPMTEIRIOSDEVGDRHRRPOEKKAKGL-----GK 170  
Db 272 GEPEPTFWNGEKSPKSGDPMTEIRIOSDEVGDRHRRPOEKKAKGLGEPRAALCEAGK 331  
  
Qy 171 EITLLMOTLNTLSTPEBKLAALCKKYAELIEHNSOKMKLQKQSQLVQEKDHLRGE 230  
Db 332 EITLLMOTLNTLSTPEBKLAALCKKYAELIEHNSOKMKLQKQSQLVQEKDHLRGE 391  
  
Qy 231 HSKAVLARSKLESICRELRNRSIK-----PEGVQARERBEKR 270  
Db 392 HSKAVLARSKLESICRELRNRSIKLGDPCVAVDDGSRADPPAQEGEVQARERBEKR 451  
  
Qy 271 KEVTSHEQVTLNDIQLQWEQHNERNRSKLQRNEMELAEKLIQVLEERIEHDKVFKHX 330  
Db 452 KEVTSHEQVTLNDIQLQWEQHNERNRSKLQRNEMELAEKLIQVLEERIEHDKVFKHX 511  
  
Qy 331 DLQOQLVDALQQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKKQALAY 390

Db 512 DLQOQLVDALQQAQEMLKEAERHOREKDFLLKEAVESQRMCELMKQOETHLKKQALAY 571  
Qy 391 TEKPFERFQNTLSKSSSEVFTTFQEMEKMTKKIKKLEKETTTMYRSRWSSNKVALLNABEK 450  
Db 572 TEKPFERFQNTLSKSSSEVFTTFQEMEKMTKKIKKLEKETTTMYRSRWSSNKVALLNABEK 631  
Qy 451 TVRDXELEGLQVKIORLEKLCRALQTERNDLNKRVDL-----SAGGQGSILT 497  
Db 632 TVRDXELEGLQVKIORLEKLCRALQTERNDLNKRVDL-----SAGGQGSILT 691  
Qy 498 DSGPERRRPEGGAQAPSS-----PRVTEAPCYPGAPSTASQGTGPQEP 541  
Db 692 BEALSRCPOEQCFMLFTSAKRNPR---KPLLPF-----DGEGLRPSQRP 733  
  
RESULT 9  
AB11764  
ID AB11764 standard; peptide; 510 AA.  
XX  
XX AB11764;  
XX AC  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human LPS binding protein homologue, SEQ ID NO:2134.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
KW chronic inflammatory condition; proliferative retinopathy;  
KW atherosclerosis; coronary heart disease; arterial ischaemia;  
KW bone disorder; osteoporosis; vascular growth disorder;  
KW tissue regeneration; wound healing; infection; immune disorder;  
KW cell culture; drug screening; gene therapy; antiinflammatory;  
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
KW antifungal; vulnery; antitumor.  
XX  
XX Homo sapiens.  
XX OS  
XX WO200157188-A2.  
XX PN  
XX  
PD 09-AUG-2001.  
XX  
XX  
PF 05-FEB-2001; 2001WO-US003800.  
XX  
XX  
PR 03-FEB-2000; 2000US-00496914.  
PR 27-APR-2000; 2000US-00560875.  
XX  
XX (HYSB-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
PI  
XX WPI; 2001-457740/49.  
XX  
XX N-PSDB; ABA09008.  
DR  
XX  
XX Human proteins and DNA encoding sequences useful for preventing, treating  
PT or ameliorating a medical condition in a mammalian subject e.g. arthritis  
PT and cancer.  
XX  
XX  
PS Claim 20; Page 240-241; 1963pp; English.  
XX  
CC Sequences AB110981-AB112330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may

CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel and  
 CC polypeptide of the invention

XX SQ Sequence 510 AA;

Query Match 86.8%; Score 2416; DB 4; Length 510;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-138;  
 Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAGKSNPKSPGQFAGPEGAQRPQAAPAVEAGPGSSQAPRKEGAQA 60  
 DB 12 MNQDKKNGAAGKSNPKSPGQFAGPEGAQRPQAAPAVEAGPGSSQAPRKEGAQA 71  
 QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNGGPGEDGAQGEPAEPDAKSRITYARN 120  
 DB 72 RTAQSGALRDVSEELSRQLEDILSTYCVNNGGPGEDGAQGEPAEPDAKSRITYARN 131  
 QY 121 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAAGLKGKETTILMOTLN 180  
 DB 132 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAAGLKGKETTILMOTLN 191  
 QY 181 TLSTPEEKLAALCKKYAEELLEHRNSQKMLLQKQSQLVQEKDHLRGHSHKAVLARSK 240  
 DB 192 TLSTPEEKLAALCKKYAEELLEHRNSQKMLLQKQSQLVQEKDHLRGHSHKAVLARSK 251  
 QY 241 LESLCRELQNRNRSLEKGVQARBEERKEKRVTSHFVTLNDIQLQMEQHNRNSKLQ 300  
 DB 252 LESLCRELQNRNRSLEKGVQARBEERKEKRVTSHFVTLNDIQLQMEQHNRNSKLQ 311  
 QY 301 ENMELAEERLKKLIROYELREEHIDKVPKHKLQQLQVDAKQQAOEMLKEAEERHQRKD 360  
 DB 312 ENMELAEERLKKLIROYELREEHIDKVPKHKLQQLQVDAKQQAOEMLKEAEERHQRKD 371  
 QY 361 FLKKAESQRMCELMKQOETHLKKQALATYRKFEFQNTLSKSEVFTTFQEMEKMTK 420  
 DB 372 FLKKAESQRMCELMKQOETHLKKQALATYRKFEFQNTLSKSEVFTTFQEMEKMTK 431  
 QY 421 KIKKLEKETTMYRSRWSSNKALLEWABEKTVRDKLEGLQVKIQRLEKLCRALQT 476  
 DB 432 KIKKLEKETTMYRSRWSSNKALLEWABEKTVRDKLEGLQVKIQRLEKLCRALQT 487

RESULT 10

AAM79741

ID AAM79741 standard; protein; 510 AA.

XX AC

XX AAM79741;

XX

DT 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 3387.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
 XX nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

XX W0200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001WO-US004098.

XX 03-FEB-2000; 2000US-00496914.

XX 21-APR-2000; 2000US-00568875.

XX 20-JUN-2000; 2000US-00598075.

XX 19-JUL-2000; 2000US-00628325.

XX 01-SEP-2000; 2000US-00654936.

XX 15-SEP-2000; 2000US-00663561.

XX 20-OCT-2000; 2000US-00693325.

XX 30-NOV-2000; 2000US-00728422.

XX (HYSB-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;

XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

XX N-PSDB; AAKS2874.

XX Nucleic acids encoding polypeptides with cytokine-like activities, useful

XX in diagnosis and gene therapy.

XX Claim 20; Page 319; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 XX encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 XX cytokine, cell proliferation or cell differentiation or which may induce  
 XX production of other cytokines in other cell populations. The  
 XX polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 XX peptide therapy. The polypeptides have various cytokine-like activities,  
 XX e.g. stem cell growth factor activity, haematopoiesis regulating  
 XX activity, tissue growth factor activity, immunomodulatory activity and  
 XX activin/inhibin activity and may be useful in the diagnosis and/or  
 XX treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 XX (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 XX sequence listing were missing at the time of publication

XX Sequence 510 AA;

Query Match 86.8%; Score 2416; DB 4; Length 510;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-138;  
 Matches 476; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNQDKKNGAAGKSNPKSPGQFAGPEGAQRPQAAPAVEAGPGSSQAPRKEGAQA 60  
 DB 12 MNQDKKNGAAGKSNPKSPGQFAGPEGAQRPQAAPAVEAGPGSSQAPRKEGAQA 71  
 QY 61 RTAQSGALRDVSEELSRQLEDILSTYCVNNGGPGEDGAQGEPAEPDAKSRITYARN 120  
 DB 72 RTAQSGALRDVSEELSRQLEDILSTYCVNNGGPGEDGAQGEPAEPDAKSRITYARN 131  
 QY 121 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAAGLKGKETTILMOTLN 180  
 DB 132 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVEGDRDHRPQEKKAAGLKGKETTILMOTLN 191  
 QY 181 TLSTPEEKLAALCKKYAEELLEHRNSQKMLLQKQSQLVQEKDHLRGHSHKAVLARSK 240

DB 192 TLSTPEKLAALCKYAELEHNSQKMLQKQSQVLQKDHGRHSHKAVLARK 251  
QY 241 LESLCRELQHNRLKEEGVQVAREEKKKVTSHFQVTLNDIQLQMEQHNRNSKLQ 300  
DB 252 LESLCRELQHNRLKEEGVQVAREEKKKVTSHFQVTLNDIQLQMEQHNRNSKLQ 311  
QY 301 ENMELAEKLIQVYELREEHIDKVPKHDLQQLQVDAKLQQAQEWLKEAERHOREKD 360  
DB 312 ENMELAEKLIQVYELREEHIDKVPKHDLQQLQVDAKLQQAQEWLKEAERHOREKD 371  
QY 361 FLKEAVESQRMCELMKQETHLQKQALYTERFEPQNTLSKSSSEVFTTFKQEMKMTX 420  
DB 372 FLKEAVESQRMCELMKQETHLQKQALYTERFEPQNTLSKSSSEVFTTFKQEMKMTX 431  
QY 421 KIKKLEKTTMYRSRWNSSNKALLEWAEKTVRDKLEGLQVQIKRLEKLCRALQ 476  
DB 432 KIKKLEKTTMYRSRWNSSNKALLEWAEKTVRDKLEGLQVQIKRLEKLCRALQ 487

RESULT 11  
ABO07223  
ID ABO07223 standard; protein; 356 AA.  
AC ABO07223;  
XX  
XX  
DT 13-AUG-2003 (first entry)  
XX  
XX Human p53 modifying protein, SEQ ID 183.  
DE  
XX Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;  
XX antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;  
XX lung cancer; ovarian cancer; angiogenesis; cell cycle;  
XX apoptotic disorder; cell proliferation disorder.  
OS Homo sapiens.  
XX  
XX WO200299122-A1.  
XX  
PD 12-DEC-2002.  
XX  
XX 03-JUN-2002; 2002WO-US017382.  
XX  
XX 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328605P.  
PR 15-FEB-2002; 2002US-0357253P.  
XX  
XX (EXEL-) EXELIXIS INC.  
XX  
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
XX  
XX WPI; 2003-156859/15.  
XX N-PSDB; ACD13448.  
XX  
XX Identifying modulators of the p53 pathway for use in treating apoptotic  
XX or cell proliferation disorders, comprises screening for agents that  
XX modulate activity of a human ortholog of genes that modify the p53  
XX pathway in Drosophila.  
XX  
XX Example 2; Page 533-534; 678pp; English.  
XX  
XX The invention relates to identifying (M1) a candidate p53 pathway  
XX modulating agent, by contacting an assay system comprising a purified HM  
XX polypeptide (human orthologue of genes that modify the p53 pathway in  
XX Drosophila) or nucleic acid with a test agent under conditions, where but  
XX for the presence of the test agent, the system provides a reference  
XX activity, and detecting a test agent-biased activity of the assay system.  
XX Also included are modulating (M2) a p53 pathway of a cell (comprising  
XX contacting a cell defective in p53 function with a candidate modulator  
XX that specifically binds to a HM polypeptide comprising an HM amino acid  
XX sequence, where p53 function is restored), modulating (M3) a p53 pathway  
XX in a mammalian cell (comprising contacting the cell with an agent that  
XX specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)

CC a disease in a patient (comprising: (a) obtaining a biological sample  
CC from the patient; (b) contacting the sample with a probe for HM  
CC expression; (c) comparing the results with a control; and (d) determining  
CC whether the comparison indicates a likelihood disease). (M1) is useful  
CC for identifying modulators of the p53 pathway. A probe for HM expression  
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
CC in a patient, where the cancer has greater than 25 % expression level.  
CC Modulators identified by (M1) are useful in a variety of diagnostic and  
CC therapeutic applications, where disease or disorder prognosis is related  
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and  
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring  
CC the p53 function of the cell, so that the cell undergoes normal  
CC proliferation or progression through the cell cycle. (M2) and (M3) are  
CC also useful for treating defects in the p53 pathway such as angiogenic,  
CC apoptotic or cell proliferation disorders. The present sequence  
CC represents a human p53 pathway modifying protein  
XX  
XX Sequence 356 AA;  
QY 191 ALCKKYAELEHNSQKMLQKQSQVLQKDHGRHSHKAVLARKSLCRLQ 250  
DB 1 ALCKKYAELEHNSQKMLQKQSQVLQKDHGRHSHKAVLARKSLCRLQ 60  
QY 251 HNRSLKEGVQVAREEKKKVTSHFQVTLNDIQLQMEQHNRNSKLQENNELAERLK 310  
DB 61 HNRSLKEGVQVAREEKKKVTSHFQVTLNDIQLQMEQHNRNSKLQENNELAERLK 120  
QY 311 KLIEQYELAREEHIDKVPKHDLQQLQVDAKLQQAQEWLKEAERHOREKDPLLKEAVESQ 370  
DB 121 KLIEQYELAREEHIDKVPKHDLQQLQVDAKLQQAQEWLKEAERHOREKDPLLKEAVESQ 180  
QY 371 RMCELMKQETHLQKQALYTERFEPQNTLSKSSSEVFTTFKQEMKMTXKIKKLEKTT 430  
DB 181 RMCELMKQETHLQKQALYTERFEPQNTLSKSSSEVFTTFKQEMKMTXKIKKLEKTT 240  
QY 431 MYRSRWNSSNKALLEWAEKTVRDKLEGLQVQIKRLEKLCRALQTERNDLNKRVQDLSA 490  
DB 241 MYRSRWNSSNKALLEWAEKTVRDKLEGLQVQIKRLEKLCRALQTERNDLNKRVQDLSA 300  
QY 491 GGQGSILTDGPERRPEGGAQAPSSPRVTEAPCYGAPSTEASGOTGPOEPTSARA 546  
DB 301 GGQGSILTDGPERRPEGGAQAPSSPRVTEAPCYGAPSTEASGOTGPOEPTSARA 356

RESULT 12  
AAM78757  
ID AAM78757 standard; protein; 358 AA.  
XX  
XX AAM78757;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
XX Human protein SEQ ID NO 1419.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorder; arthritis; inflammation.  
XX Homo sapiens.  
XX  
XX WO200157190-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US004098.  
XX  
XX 03-FEB-2000; 2000US-00496914.  
PR

RESULT 13  
AAB93594  
ID AAB9  
XX  
AC AAB9  
XX



CC also useful for treating defects in the p53 pathway such as angiogenic,  
CC apoptotic or cell proliferation disorders. The present sequence  
CC represents a human p53 pathway modifying protein  
XX  
SQ Sequence 395 AA;

Query Match 45.4%; Score 1265.5; DB 6; Length 395;  
Best Local Similarity 65.2%; Pred. No. 1.9e-68;  
Matches 255; Conservative 59; Mismatches 50; Indels 27; Gaps 5;

QY 169 GKEITLIMQTLNTLSTPEEKLAALCKKYAELLBEHNSOKMGLKQKQOLVQEKDHLR 228  
DB 3 GKEVLIMQALNTLSTPEEKLAALCKKYADLLSESVQKMKILQKQAOIVKEKVLHQ 62  
QY 229 GEHSAVLAEKLSLRELRNRSIKKEGVORAREBEERKEVTSHFVOTLNDIOLQM 288  
DB 63 SEHSAVLAEKLSLRELRNRSIKKEGVORAREBEERKEVTSHFVOTLNDIOLQM 122  
QY 289 EQHNSKLRQRNMLAEKLLKLIQYELREHIDKVFHKLQOLVQAKLQQAQEML 348  
DB 123 EQHNSKLRQRNMLAEKLLKLIQYELREHIDKVFHKLQOLVQAKLQQAQEML 182  
QY 349 KEAEHREKDFLLKEAVESQRMCELMKQOETHLQOOLALYTEKEFEFQNTLSKSEVF 408  
DB 183 KEAEHREKDFLLKEAVESQRMCELMKQOETHLQOOLALYTEKEFEFQNTLSKSEVF 242  
QY 409 TTFKQEMKMTKKIKKLEKETTMYRSWESSNALLMABEKTVRDKELGLQVKIQRLE 468  
DB 243 TTFKQEMKMTKKIKKLEKETTMYRSWESSNALLMABEKTVRDKELGLQVKIQRLE 302  
QY 469 KLCRALQTERNDLKNRVQ-----DLGAGGQSLT--DSGPE-----RR 504  
DB 303 KLCRALQTERNDLKNRVQ-----DLGAGGQSLT--DSGPE-----RR 504  
QY 505 PEGPGAQA-PSSPR--VTEAPCYFGAPSTEA 532  
DB 363 ALGAHLEAEPKSORSQSAVQKPPSTGSAPIES 393

Search completed: June 8, 2004, 16:32:54  
Job time : 62 secs

QY 541 PTSA 544  
DB 574 PSKA 577

RESULT 15  
ABC07222 standard; protein; 395 AA.

XX ABC07222;

XX 13-AUG-2003 (first entry)

XX Human p53 modifying protein, SEQ ID 182.

XX Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;  
XX antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;  
XX lung cancer; ovarian cancer; angiogenesis; cell cycle;  
XX apoptotic disorder; cell proliferation disorder.

XX Homo sapiens.

XX WO200299122-A1.

XX 12-DEC-2002.

XX 03-JUN-2002; 2002WO-US017382.

XX 05-JUN-2001; 2001US-0296076P.

XX 10-OCT-2001; 2001US-0328605P.

XX 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-156859/15.

XX N-PSDB; ACD13396.

XX Identifying modulators of the p53 pathway for use in treating apoptotic  
XX or cell proliferation disorders, comprises screening for agents that  
XX modulate activity of a human ortholog of genes that modify the p53  
XX pathway in Drosophila.

XX Example 2: Page 532-533; 678pp; English.

XX The invention relates to identifying (M1) a candidate p53 pathway  
XX modulating agent, by contacting an assay system comprising a purified HM  
XX polypeptide (human ortholog of genes that modify the p53 pathway in  
XX Drosophila) or nucleic acid with a test agent under conditions, where but  
XX for the presence of the test agent, the system provides a reference  
XX activity, and detecting a test agent-biased activity of the assay system.  
XX Also included are modulating (M2) a p53 pathway of a cell (comprising  
XX contacting a cell defective in p53 function with a candidate modulator  
XX that specifically binds to a HM polypeptide comprising an HM amino acid  
XX sequence, where p53 function is restored), modulating (M3) a p53 pathway  
XX in a mammalian cell (comprising contacting the cell with an agent that  
XX specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
XX a disease in a patient (comprising: (a) obtaining a biological sample  
XX from the patient; (b) contacting the sample with a probe for HM  
XX expression; (c) comparing the results with a control; and (d) determining  
XX whether the comparison indicates a likelihood disease). (M1) is useful  
XX for identifying modulators of the p53 pathway. A probe for HM expression  
XX is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
XX in a patient, where the cancer has greater than 25 % expression level.  
XX Modulators identified by (M1) are useful in a variety of diagnostic and  
XX therapeutic applications, where disease or disorder prognosis is related  
XX to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
XX proliferation disorders (e.g. cancer). Another two new methods (M2 and  
XX M3) are useful for modulating the p53 pathway of a cell, thus restoring  
XX the p53 function of the cell, so that the cell undergoes normal  
XX proliferation or progression through the cell cycle. (M2) and (M3) are

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:33:01 ; Search time 23 Seconds  
(without alignments)  
1225.556 Million cell updates/sec

Title: US-10-023-529-44  
Perfect score: 2785  
Sequence: 1 MNQDKKNGAKQSNPKSSP.....APSTEASGQTGPQEPSTASARA 546

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgm2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgm2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgm2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgm2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgm2\_6/ptodata/2/iaa/PCTUS COXA.pep.\*  
6: /cgm2\_6/ptodata/2/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2785	100.0	546	4	US-09-616-289-44
2	2693	95.7	530	4	US-08-979-608A-8
3	2693	96.7	530	4	US-09-517-849-8
4	2693	96.7	530	4	US-09-616-289-8
5	2552.5	91.7	557	4	US-08-979-608A-5
6	2552.5	91.7	557	4	US-09-517-849-5
7	2552.5	91.7	557	4	US-09-616-289-5
8	1039.5	37.3	386	2	US-08-968-751-6
9	252	9.0	787	4	US-09-688-188B-151
10	252	9.0	787	4	US-09-291-417D-151
11	251.5	9.0	786	4	US-09-688-188B-23
12	251.5	9.0	786	4	US-09-291-417D-23
13	251.5	9.0	1001	4	US-09-688-188B-31
14	251.5	9.0	1001	4	US-09-291-417D-31
15	250	9.0	1001	3	US-09-060-410-2
16	250	9.0	1001	4	US-09-723-458-2
17	243	8.7	1939	4	US-09-310-187A-1
18	242.5	8.7	1898	1	US-08-056-200-94
19	242.5	8.7	1898	4	US-08-800-644-94
20	240	8.6	1162	2	US-08-728-323A-2
21	240	8.6	1162	4	US-09-298-568-2
22	240	8.6	1162	4	US-09-410-399-2
23	238	8.5	2662	4	US-09-595-684B-31
24	235	8.4	1886	4	US-08-938-105-3
25	229.5	8.2	3878	4	US-09-914-259-11
26	225.5	8.1	3248	1	US-08-353-700-1
27	225.5	8.1	3248	5	PCT-US95-16216-1

ALIGNMENTS

RESULT 1

US-09-616-289-44

; Sequence 44, Application US/09616289  
; Patent No. 6632923  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/09/616,289  
; CURRENT FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 09/517,849  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; PRIOR FILING DATE: 1997-06-03  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 44  
; LENGTH: 546  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-616-289-44

Query Match 100.0%; Score 2785; DB 4; Length 546;  
Best Local Similarity 100.0%; Pred. No. 1.2e-178;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNQDKKNGAKQSNPKSSPGQEPAGEGQERPSQAAPAVEAEGSGSSQAPRKEGAQA	60
Db	1	MNQDKKNGAKQSNPKSSPGQEPAGEGQERPSQAAPAVEAEGSGSSQAPRKEGAQA	60
Qy	61	RTAQSGALRDVSELSRQLEDILSTYCDVNNQSGPDGAGQEPAPEDAEKSRITYVARN	120
Db	61	RTAQSGALRDVSELSRQLEDILSTYCDVNNQSGPDGAGQEPAPEDAEKSRITYVARN	120
Qy	121	GEPEPTPVNGKEPKSGKDPNTEIROSDVGRDHRHPOEKKAKGLKEITLLMQTLN	180
Db	121	GEPEPTPVNGKEPKSGKDPNTEIROSDVGRDHRHPOEKKAKGLKEITLLMQTLN	180
Qy	181	TLSTPEEKLAALCKKYAEELLEHRNSQMKLLQKQSQLVQEKDHLRGHSHKAVLARSK	240
Db	181	TLSTPEEKLAALCKKYAEELLEHRNSQMKLLQKQSQLVQEKDHLRGHSHKAVLARSK	240



Qy 241 LESLCRLQNRSLKEGVQVRAEEBKKRKEVTSHTFQVTLNDIQLQMEQHNRSLRQ 300  
Db 241 LESLCRLQNRSLKEGVQVRAEEBKKRKEVTSHTFQVTLNDIQLQMEQHNRSLRQ 300  
Qy 301 ENMELARLKLIOYELREBHDVFKHDLQQLVDKIQQAQCEMLKEAEERHOREKD 360  
Db 301 ENMELARLKLIOYELREBHDVFKHDLQQLVDKIQQAQCEMLKEAEERHOREKD 360  
Qy 361 FLLKEAVESQRMCLMKQOETHLQKQALYTEKEPFBFQNTLSKSSEVFTTFKQMEKMTK 420  
Db 361 FLLKEAVESQRMCLMKQOETHLQKQALYTEKEPFBFQNTLSKSSEVFTTFKQMEKMTK 420  
Qy 421 KIKLEKETTMYRWRSSNKLLEMAEKTVRDKLEGLQVKIORLEKLCRALQTERND 480  
Db 421 KIKLEKETTMYRWRSSNKLLEMAEKTVRDKLEGLQVKIORLEKLCRALQTERND 480  
Qy 481 LNKRVQDLASQGGSLTDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTASQGTGPOE 540  
Db 481 LNKRVQDLASQGGSLTDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTASQGTGPOE 540  
Qy 541 PTSARA 546  
Db 541 PTSARA 546

RESULT 2  
US-08-979-608A-8  
; Sequence 8, Application US/08979608A  
; Patent No. 6355451  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; Lees, Robert S.  
; Law, Simon W.  
; Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
; TREATING ATHEROSCLEROSIS  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/979,608A  
; FILING DATE: 26-NOV-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/048,547  
; FILING DATE: 03-JUN-1997  
; APPLICATION NUMBER: US 60/031,930  
; FILING DATE: 27-NOV-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Louis  
; REGISTRATION NUMBER: 35,965  
; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 530 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-08-979-608A-8

Query Match 96.7%; Score 2693; DB 4; Length 530;  
Best Local Similarity 99.8%; Pred. No. 1.6e-172; Indels 0; Gaps 0;  
Matches 529; Conservative 0; Mismatches 1;  
Qy 17 KSPGQPEAGPEGAQERPQAAAPAVEAEGFGSSQAPRKEGAQARTAQSGALRDVSEELS 76  
Db 1 KSPGQPEAGPEGAQERPQAAAPAVEAEGFGSSQAPRKEGAQARTAQSGALRDVSEELS 60  
Qy 77 RQLEDLSTYCVDNNGGGEDEGAQEPAPEDAKSRITYVARNGEPETPTVVNGKEPS 136  
Db 61 RQLEDLSTYCVDNNGGGEDEGAQEPAPEDAKSRITYVARNGEPETPTVVNGKEPS 120  
Qy 137 KGPNTTEEIRQSDDEVDRDRRPFQEKCKAGLKGKXITLLMQTLNLTSTPEEKLAALCKKY 196  
Db 121 KGPNTTEEIRQSDDEVDRDRRPFQEKCKAGLKGKXITLLMQTLNLTSTPEEKLAALCKKY 180  
Qy 197 AELEBHRNSQOMKLLQKQSLVQEKDHLRGEHSKAVLARSKLESICRELOHNRSLK 256  
Db 181 AELEBHRNSQOMKLLQKQSLVQEKDHLRGEHSKAVLARSKLESICRELOHNRSLK 240  
Qy 257 EGVQVRAEEBKKRKEVTSHTFQVTLNDIQLQMEQHNRSLRQENMELARLKLIOY 316  
Db 241 EGVQVRAEEBKKRKEVTSHTFQVTLNDIQLQMEQHNRSLRQENMELARLKLIOY 300  
Qy 317 ELREHIDKVFHKDLQQLVDKIQQAQCEMLKEAEERHOREKD FLLKEAVESQRMCELM 376  
Db 301 ELREHIDKVFHKDLQQLVDKIQQAQCEMLKEAEERHOREKD FLLKEAVESQRMCELM 360  
Qy 377 KQOETHLKQALYTEKEPFBFQNTLSKSSEVFTTFKQMEKMTKIKLEKETTYRWSRW 436  
Db 361 KQOETHLKQALYTEKEPFBFQNTLSKSSEVFTTFKQMEKMTKIKLEKETTYRWSRW 420  
Qy 437 ESSNKALLEMAEKTVRDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLASQGGSL 496  
Db 421 ESSNKALLEMAEKTVRDKLEGLQVKIORLEKLCRALQTERNDLNKRVQDLASQGGSL 480  
Qy 497 TDSGPRRPEGGAQAPSSPRVTEAPCYPCAPSTASQGTGPOEPTSARA 546  
Db 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPCAPSTASQGTGPOEPTSARA 530

RESULT 3  
US-09-517-849-8  
; Sequence 8, Application US/09517849  
; Patent No. 6605588  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; Lees, Robert S.  
; Law, Simon W.  
; Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
; TREATING ATHEROSCLEROSIS  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/517,849  
; FILING DATE: 02-Mar-2000  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/979,608

FILING DATE: 26-NOV-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis  
REGISTRATION NUMBER: 35,955  
REFERENCE/DOCKET NUMBER: 10797-003001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-8070  
TELEFAX: 617/542-8906  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 530 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-517-849-8

Query Match 96.7%; Score 2693; DB 4; Length 530;  
Best Local Similarity 99.8%; Pred. No. 1.6e-172;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 76  
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 60  
QY 77 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPAEDAEKSRITYVARNGEPPTPVVNGEKEPS 136  
DB 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPAEDAEKSRITYVARNGEPPTPVVNGEKEPS 120  
QY 137 KGPNTETIRQSDVGDHRHRRPOEKKAGLKEITLLMOTLNTLSTPEEKLAALCKKY 196  
DB 121 KGPNTETIRQSDVGDHRHRRPOEKKAGLKEITLLMOTLNTLSTPEEKLAALCKKY 180  
QY 197 AELLERHNSQOMKLLQKQSQVLQVQKDLHRLGEHSAVLARSKLESICRELQRHNSLK 256  
DB 181 AELLERHNSQOMKLLQKQSQVLQVQKDLHRLGEHSAVLARSKLESICRELQRHNSLK 240  
QY 257 EBGVQARBEERKEKVTSHFOVTLNDIQLQMEQHNRNSKLRQNMELAEKLLIEQY 316  
DB 241 EBGVQARBEERKEKVTSHFOVTLNDIQLQMEQHNRNSKLRQNMELAEKLLIEQY 300  
QY 317 ELREHIDKVFHKDLQQLVDKALQQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 376  
DB 301 ELREHIDKVFHKDLQQLVDKALQQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 360  
QY 377 KQETHLKKQALALYTEKFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKTTMYRSRW 436  
DB 361 KQETHLKKQALALYTEKFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKTTMYRSRW 420  
QY 437 ESSNKALLEMAEKTVRDKELEGLOVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLL 496  
DB 421 ESSNKALLEMAEKTVRDKELEGLOVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLL 480  
QY 497 TDSGPERRPGGAQAPSSPRVTAPCPYGPAPSTASGQTGPQEPPTSARA 546  
DB 481 TDSGPERRPGGAQAPSSPRVTAPCPYGPAPSTASGQTGPQEPPTSARA 530

RESULT 5  
US-08-979-608A-5  
Sequence 5, Application US/08979608A  
Patent No. 6355451  
GENERAL INFORMATION:  
APPLICANT: Lees, Ann M.  
Lees, Robert S.  
Law, Simon W.  
Arjona, Anibal A.  
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
TREATING ATHEROSCLEROSIS  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street

Query Match 96.7%; Score 2693; DB 4; Length 530;  
Best Local Similarity 99.8%; Pred. No. 1.6e-172;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 17 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 76  
DB 1 KSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAAPKPEGAQARTAQSGALRDVSEELS 60  
QY 77 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPAEDAEKSRITYVARNGEPPTPVVNGEKEPS 136  
DB 61 RQLEDILSTYCVNNQGGPGEDGAQGEPAEPAEDAEKSRITYVARNGEPPTPVVNGEKEPS 120  
QY 137 KGPNTETIRQSDVGDHRHRRPOEKKAGLKEITLLMOTLNTLSTPEEKLAALCKKY 196  
DB 121 KGPNTETIRQSDVGDHRHRRPOEKKAGLKEITLLMOTLNTLSTPEEKLAALCKKY 180  
QY 197 AELLERHNSQOMKLLQKQSQVLQVQKDLHRLGEHSAVLARSKLESICRELQRHNSLK 256  
DB 181 AELLERHNSQOMKLLQKQSQVLQVQKDLHRLGEHSAVLARSKLESICRELQRHNSLK 240  
QY 257 EBGVQARBEERKEKVTSHFOVTLNDIQLQMEQHNRNSKLRQNMELAEKLLIEQY 316  
DB 241 EBGVQARBEERKEKVTSHFOVTLNDIQLQMEQHNRNSKLRQNMELAEKLLIEQY 300  
QY 317 ELREHIDKVFHKDLQQLVDKALQQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 376  
DB 301 ELREHIDKVFHKDLQQLVDKALQQAQEMLEAEERHOREKDFLLKEAVESQRMCELM 360  
QY 377 KQETHLKKQALALYTEKFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKTTMYRSRW 436  
DB 361 KQETHLKKQALALYTEKFEFQNTLSKSEVFTTFKQEMEKMTKKIKKLEKTTMYRSRW 420  
QY 437 ESSNKALLEMAEKTVRDKELEGLOVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLL 496  
DB 421 ESSNKALLEMAEKTVRDKELEGLOVKIQRLKLCRALQTERNDLNKRVQDLSAGQGSLL 480  
QY 497 TDSGPERRPGGAQAPSSPRVTAPCPYGPAPSTASGQTGPQEPPTSARA 546  
DB 481 TDSGPERRPGGAQAPSSPRVTAPCPYGPAPSTASGQTGPQEPPTSARA 530

RESULT 4  
US-09-516-289-8  
Sequence 8, Application US/09616289  
Patent No. 6632923  
GENERAL INFORMATION:  
APPLICANT: Lees, Ann M.  
Lees, Robert S.  
Law, Simon W.  
Arjona, Anibal A.  
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
ATHEROSCLEROSIS  
FILE REFERENCE: 10797-004001  
CURRENT APPLICATION NUMBER: US/09/616,289  
CURRENT FILING DATE: 2000-07-14



QY 240 KLESCLRELQHRNRSLEKEGVQARAREEKEKVTSHFQVTLNDIOLQMEQHNRNSKLR 239  
DB 240 KLESCLRELQHRNRSLEKEGVQARAREEKEKVTSHFQVTLNDIOLQMEQHNRNSKLR 239  
QY 300 QENWELAEKRLKLEIQEYELREEHIDKVFHKHKLQOQQLVDKALQQAQEMLEKAEERHOREK 359  
DB 300 QENWELAEKRLKLEIQEYELREEHIDKVFHKHKLQOQQLVDKALQQAQEMLEKAEERHOREK 359  
QY 360 DFLKKEAVESQRMCELMKQOETHLKOQALYTEKFEFQNTLSKSEVFTTFKQEMKMT 419  
DB 360 DFLKKEAVESQRMCELMKQOETHLKOQALYTEKFEFQNTLSKSEVFTTFKQEMKMT 419  
QY 420 KIKKLEKETTTMYRSRWESSNKALLEMAEKTLDKLEGLQVKIQRLEKLCALQOTERN 479  
DB 420 KIKKLEKETTTMYRSRWESSNKALLEMAEKTLDKLEGLQVKIQRLEKLCALQOTERN 479  
QY 480 DLNKRVDLSAGQGSULTDGGPRRP-----EGGGAQAPSSPRVTAPCYPGAP 528  
DB 480 DLNKRVDLSAGQGSULTDGGPRRP-----EGGGAQAPSSPRVTAPCYPGAP 528  
QY 529 STEASGOTGPQEPPTATA 546  
DB 540 STEASGOTGPQEPPTATA 557

RESULT 7  
US-09-616-289-5  
; Sequence 5, Application US/09616289  
; Patent No. 6632923  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/09/616,289  
; CURRENT FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 09/517,849  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; PRIOR FILING DATE: 1997-06-03  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 557  
; TYPE: PRT  
; ORGANISM: Oryctolagus cuniculus  
US-09-616-289-5

Query Match 91.7%; Score 2552.5; DB 4; Length 557;  
Best Local Similarity 91.0%; Pred. No. 4.4e-183;  
Matches 508; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

QY 1 MKNQDKNGAQNPKSSPCQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKEGAQA 60  
DB 1 MKNQDKNGAQNPKSSPCQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKEGAQA 59  
QY 61 RTAQSGALRDVSELSQLEIDILSTYCVDNQGGPDGAGQGPAPAEPAEDAEKRTYVARN 120  
DB 60 RTAQSGALRDVSELSQLEIDILSTYCVDNQGGPDGAGQGPAPAEPAEDAEKRTYVARN 119  
QY 121 GEPEP-TPVNVGRKEPKSGDNPTEIRQSEVGVGDHRRPOEKKAAGLGEKITLLMOTL 179  
DB 120 GEPEP-TPVNVGRKEPKSGDNPTEIRQSEVGVGDHRRPOEKKAAGLGEKITLLMOTL 179

QY 180 NTLSTPEEKLAALCKKYAELLBEHRNSOKOMKLLQKOSQOLVQEKDHLRGEHSAVLARS 239  
DB 180 NTLSTPEEKLAALCKKYAELLBEHRNSOKOMKLLQKOSQOLVQEKDHLRGEHSAVLARS 239  
QY 240 KLESCLRELQHRNRSLEKEGVQARAREEKEKVTSHFQVTLNDIOLQMEQHNRNSKLR 239  
DB 240 KLESCLRELQHRNRSLEKEGVQARAREEKEKVTSHFQVTLNDIOLQMEQHNRNSKLR 239  
QY 300 QENWELAEKRLKLEIQEYELREEHIDKVFHKHKLQOQQLVDKALQQAQEMLEKAEERHOREK 359  
DB 300 QENWELAEKRLKLEIQEYELREEHIDKVFHKHKLQOQQLVDKALQQAQEMLEKAEERHOREK 359  
QY 360 DFLKKEAVESQRMCELMKQOETHLKOQALYTEKFEFQNTLSKSEVFTTFKQEMKMT 419  
DB 360 DFLKKEAVESQRMCELMKQOETHLKOQALYTEKFEFQNTLSKSEVFTTFKQEMKMT 419  
QY 420 KIKKLEKETTTMYRSRWESSNKALLEMAEKTLDKLEGLQVKIQRLEKLCALQOTERN 479  
DB 420 KIKKLEKETTTMYRSRWESSNKALLEMAEKTLDKLEGLQVKIQRLEKLCALQOTERN 479  
QY 480 DLNKRVDLSAGQGSULTDGGPRRP-----EGGGAQAPSSPRVTAPCYPGAP 528  
DB 480 DLNKRVDLSAGQGSULTDGGPRRP-----EGGGAQAPSSPRVTAPCYPGAP 528  
QY 529 STEASGOTGPQEPPTATA 546  
DB 540 STEASGOTGPQEPPTATA 557

RESULT 8  
US-08-968-751-6  
; Sequence 6, Application US/08968751  
; Patent No. 5948643  
; GENERAL INFORMATION:  
; APPLICANT: Rubinfeld, Bonnie  
; APPLICANT: Polakis, Paul G.  
; APPLICANT: Ligandfelter, Carol  
; APPLICANT: Vuong, Terilyn T.  
; TITLE OF INVENTION: MODULATORS OF BRCA1 ACTIVITY  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ONYX Pharmaceuticals, Inc.  
; STREET: 3031 Research Drive  
; CITY: Richmond  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94806  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968,751  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Giotta, Gregory  
; REGISTRATION NUMBER: 32,028  
; REFERENCE/DOCKET NUMBER: ONYX1024 GG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 262-8710  
; TELEFAX: (510) 222-9758  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 386 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-968-751-6

Query Match 37.3%; Score 1039.5; DB 2; Length 386;  
Best Local Similarity 56.0%; Pred. No. 4.8e-62;

```
Matches 225; Conservative 55; Mismatches 99; Indels 23; Gaps 6;
QY 29 GAQRPQQAAPAVEAGPGSSQAARK--PEGAQAATAQSGALRDVSBELSGQLEDILSTY 86
DB 4 GAEE-----ATEAGRGRRRSRQKPEIGTMBEAGICGLGVKADMLCNSQSNLILQHQ 56
QY 87 CVDNNOGPGEDGAGGAPAEADAESRTYVARNGEPPTPVNGEKPSKGDPTNTEIR 146
DB 57 --GSMCGTSNKHSLSEDEGSDPTENNLVS-----PAYCTQESREIPEG--EAR 104
QY 147 QSDVGDGRDHRPQPKKXKARGIKREITLLMOTLNTLSTPEEKLAALCKKYAELLEHRNS 206
DB 105 TDPDQGDQSECNRNKEKT--LGKVLLEMLQALNTLSTPEEKLAALCKKYADLLEERSV 162
QY 207 QOKMKLQKQSOVLQOEKHLRGHESKAVLARSKLESICRLQSRHNSLKEGVQARAE 266
DB 163 QOKMKLQKQQAIVKEKVLHQSESKAILARSKLESICRLQSRHNSLKEGVQARAE 222
QY 267 EEKKEVTSHTFQVTLNDIQLQMEQHNRNSKLRQENMELARLKLIEQYELREHIDKY 326
DB 223 EERRTEATAHQITLNEIQALQEQHDIHNAKLQENIELGEKAKLIEQYALREHIDKY 282
QY 327 FKHKLOOQVDAKLOOQAEMLKEABERHOREKDFLLKEAVESQRMCELMKQOETHLQO 386
DB 283 FKHKLOOQVDAKLOOQVTLQIKEADEKHOREBEPFLKKEATESRHKYEQMKQOQVQLKQ 342
QY 387 LALYTEKPEEFONTLSKSEVFTTTPKQEMKQVTKIKKLEKE 428
DB 343 LSLYMDKKEEFQTTAKSNELFTTTPQEMKQVTKIKKXKXK 384

RESULT 9
US-09-688-188B-151
; Sequence 151, Application US/09688188B
; Patent No. 6656716
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0328
; CURRENT APPLICATION NUMBER: US/09/688,188B
; CURRENT FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 09/291,417
; PRIOR FILING DATE: 1999-04-14
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 787
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-688-188B-151

Query Match 9.0%; Score 252; DB 4; Length 787;
Best Local Similarity 21.1%; Pred. No. 5.7e-09;
Matches 140; Conservative 121; Mismatches 215; Indels 186; Gaps 29;
QY 4 QDKKNGAAKSNPKSPGQPEAGPEGQAQRPQQAAPAVEAGPGSSQA-PRKPEGAQART 62
DB 104 QEAHNGPAYEAQEEEEQDHGVGRTGT-----VNSVGSNQSIPSMSISASSQS 151
QY 63 AQSGALRDVSBELSGQLEDILSTYCVDDNQ-----GPGEDG--AQGEPAEPEDAESRTY 116
DB 152 SSVNSLPDVSDKKS-ELDMMEGDHTVMSNSVTHLKPEENYREGDP-----RT- 200
QY 117 VARNGEPPTPVVNGEKPSKGDPTNTEIRQSDVGDGRDHRPQPKKXKAGLKGIT--- 173
DB 201 --RASDPQGPQVSRHKSHYRNREHFATIRTASLV---TROMQEHQDSLEQMSGYK 254
QY 174 -LLMOTLNTLSTPEEKLAALCKKYA---ELLEENH--SQOKMKLQKQSOQLQVE--- 223
```

```
DB 255 RMRQHQKQLMTLENKLAEMDEHRLRLDKCLTGTGRNNFAAEMEXLKKHQAAMEKAV 314
QY 224 -----KDHLEGEHSAVLARSKLESICRLQSRHNSLKEGVOR-----AREEE 268
DB 315 MSNEEKFKQOHIQAQOKKEL--NSFLESQKREYKLRKQELKEELNENQSTPKKBEQWLS 372
QY 269 KRKEVTSHTFOV--TLNDIOLQME-----QHN-----ERNSKLRQEN 302
DB 373 KQENLCHQQAEBEANLLRQCYLECECFKERMMLGHEHLEQDLVREELNKRQTKD 432
QY 303 MELAERLKLLIEQYELREHIDKVFHKHLOLQOOLVDAKLOOQAEMLKEABERHOREKDFL 362
DB 433 LEHAMLRLQHESMQEELFRHLNTIQK--MRCELI--RLQHQTELNTQLEYNKRERELR 487
QY 363 LKEAVESQRMCELMKQOETHLQOOL-----ALYTEKPEEFONTLSKSEVFTTTPKQEMK 418
DB 488 RKHVMVRQPKSLKSKELQIKKQFQDTCKIQTRQYKALRNHLE-----TTKSEHKA 542
QY 419 TKKIKKLEKETMYRGRWESSNKALLEMAEKTVR----- 453
DB 543 LKRLK--EETRLKAILABQYDHSINEMLSLSTQALRLDEAQAECQVLMKLOQELBLNA 600
QY 454 -----DKLEGLQVKI-----ORLEKLCRALQTERND-----LNKRQVD 487
DB 601 YQSKIKNQAQAQHDRLRELEQRVSLRRALLEOKIIEMLALQNERTERIRSLERQARE 660
QY 488 LSAGQGS-----LTDSGFERPRPGGAQAPSPRVTEAPCTYPGAP--STEASQOT 536
DB 661 IEAPDSGSMELGFSNMVLSNLSPE-----AFSHSYPGASGWSHNP TGSP 704
QY 537 GP 538
DB 705 GP 706

RESULT 10
US-09-291-417D-151
; Sequence 151, Application US/09291417D
; Patent No. 6680170
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT APPLICATION NUMBER: US/09/291,417D
; CURRENT FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 151
; LENGTH: 787
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-291-417D-151

Query Match 9.0%; Score 252; DB 4; Length 787;
Best Local Similarity 21.1%; Pred. No. 5.7e-09;
Matches 140; Conservative 121; Mismatches 215; Indels 186; Gaps 29;
QY 4 QDKKNGAAKSNPKSPGQPEAGPEGQAQRPQQAAPAVEAGPGSSQA-PRKPEGAQART 62
DB 104 QEAHNGPAYEAQEEEEQDHGVGRTGT-----VNSVGSNQSIPSMSISASSQS 151
QY 63 AQSGALRDVSBELSGQLEDILSTYCVDDNQ-----GPGEDG--AQGEPAEPEDAESRTY 116
DB 152 SSVNSLPDVSDKKS-ELDMMEGDHTVMSNSVTHLKPEENYREGDP-----RT- 200
QY 117 VARNGEPPTPVVNGEKPSKGDPTNTEIRQSDVGDGRDHRPQPKKXKAGLKGIT--- 173
DB 201 --RASDPQGPQVSRHKSHYRNREHFATIRTASLV---TROMQEHQDSLEQMSGYK 254
```

QY 174 -LIMQTLNTLSTPEEKLAALCKKYYA-----ELLEHN--SOKMKLLOKQSOLOVBE---223  
DB 255 RMRQHCKQLMTLENLKAEMDEHRLRLDKLETGNNFAAEMEKLIKQQAAMEKEAKV 314  
QY 224 -----KDHURGEHSAVLARSKLESCLRELORHNSLKKEGVOR-----ARBEER---268  
DB 315 NGNEEKXFOOHIQAQCKKEL--NSFLESQKREYKLRKQOLKBELENENQSTPKKEKOEMLS 372  
QY 269 KRKEVTSHPQV--TLNDIOLQME-----OHN-----ERNSKLRQENMELARLKKLIEQVELRE 302  
DB 373 KQKENIQHQAEBEAMLRQRYLECECRPKRMLGRHNLQDVLREELNKRQTKQD 432  
QY 303 MELAEKRLKLIBOYELREHIDKVFHKLQLOOLVDAKLQQAQOEMKBAEERHOREKDFL 362  
DB 433 LEHAMLARQESHQOELEFRHLATIQK--MRCELI--RLQHOTELTNQLEYNKRERELR 487  
QY 363 LKEAVESORCMELKQOETHLKOOL---ALYTEKEPEFQNTLSKSEVFTTFKQEMECM 418  
DB 488 RHMVMEVRQPSKLSKELQIKQFODTKCIQTRYKALRNHLE-----TPPKSEHKA 542  
QY 419 TKKIKKLEKETTMYRSWSSNKALLEMAEKTVR-----453  
DB 543 LKRLK--EQTRKLAIALAEQYDHSINEMLSLQALRLDEAQAECQVLKQMLQOQLELELNA 600  
QY 454 -----DKELEGLQVKI-----ORLEKLCRALQTERND-----LNKRQVOD 487  
DB 601 YQSKIQAQAQADREHLEQVLSRRALLQEKIEEMMALQNERTERIRSLERQARE 660  
QY 488 LSAGGQGS-----LTDGSPERPPEGGAQAPSSPRVTEAPCYGAP--STASGOT 536  
DB 661 IBAPDSMRLGFSNMVLSNLSPE-----AFSHSYPGASGSHNFTGGP 704  
QY 537 GP 538  
DB 705 GP 706

RESULT 11  
US-09-688-188B-23  
; Sequence 23, Application US/09688188B  
; Patent No. 6656716  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: MARTINEZ, RICARDO  
; APPLICANT: WHYTE, DAVID  
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES  
; FILE REFERENCE: 038602/0328  
; CURRENT APPLICATION NUMBER: US/09/688,188B  
; PRIOR FILING DATE: 2000-10-16  
; PRIOR FILING DATE: 09/291,417  
; PRIOR FILING DATE: 1999-04-14  
; PRIOR FILING DATE: 60/081,784  
; PRIOR FILING DATE: 1998-04-14  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 786  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-688-188B-23

Query Match 9.0%; Score 251.5; DB 4; Length 786;  
Best Local Similarity 21.6%; Pred. No. 6.1e-09;  
Matches 139; Conservative 118; Mismatches 196; Indels 191; Gaps 29;  
QY 39 PAVAE-----GPGSSQA-PRKEGAQAARTAQSGALRDVSELSRQLE 80  
DB 109 PAVAEQEEBQDHGVRTVNSVGNQSIPEMSISASSQSSVNSLPDVSDDKS-ELD 167  
QY 81 DILSTYCVDNNGQ-----GPGEDG--AQGEPAEPDAEKSRITYVARNGEPEFTPVNGEKE 134  
DB 168 MMEGDHTWMSNSVHLKPEENYREBGP-----RT---RASDPQSPQVSRHKS 215

QY 135 PSKGDPTTEIROQSDVGDRRHRRPOEKKKAGLGEKIT-----LIMQTLNTLSTPEEKLA 190  
DB 216 HYRNREHPATIRASLV-----TRQWQHEHQDSELEEQMSGYKRMERQHQKQMLTLENLKL 271  
QY 191 ALCKKYYA-----ELLEHN--SOKMKLLOKQSOLOVBE-----KDHURGEHSA 233  
DB 272 AKMEHRLRLDKLETGNNFAAEMEKLIKQQAAMEKEAKVMSNEEKKFOOHIQAQCKK 331  
QY 234 AVLARSKLESCLRELORHNSLKKEGVOR-----ARBEER---KRKEVTSHPQV--TLNDI 284  
DB 332 EL--NSFLESQKREYKLRKQOLKBELENENQSTPKKEKOEMLSQKENTQHTQAEBEAMLR 389  
QY 285 OLQME-----OHN-----ERNSKLRQENMELARLKKLIEQVELRE 320  
DB 336 RRQRYLECECRPKRMLGRHNLQDVLREELNKRQTKQDLEHAMLARQESHQOELEF 449  
QY 321 EHDKVFHKLQLOOLVDAKLQQAQOEMKBAEERHOREKDFLKEAVESORCMELKQOE 380  
DB 450 RLHNTIQK--MRCELI--RLQHOTELTNQLEYNKRERELRKHVMEVRQPSKLSKE 504  
QY 381 THLKOOL---ALYTEKEPEFQNTLSKSEVFTTFKQEMEKTKIKLEKETMYRSRW 436  
DB 505 LQIKKQFODTKCIQTRYKALRNHLE-----TPPKSEHKA 557  
QY 437 ESSNKALLEMAEKTVR-----DKELE 458  
DB 558 EYDHSINEMLSLQALRLDEAQAECQVLKQMLQOQLELLNAYQSKIRMQAQAQADREH 617  
QY 459 GLQVKI-----ORLEKLCRALQTERND-----LNKRQVODLSAGGQGS-----L 496  
DB 618 ELEQVLSRRALLQEKIEEMMALQNERTERIRSLERQAREIEAFDSMRLGFSNMVL 677  
QY 497 TDSGPERPPEGGAQAPSSPRVTEAPCYGAP--STASGOTGP 538  
DB 678 SNLSPE-----AFSHSYPGASGSHNFTGGP 705

RESULT 12  
US-09-291-417D-23  
; Sequence 23, Application US/09291417D  
; Patent No. 6680170  
; GENERAL INFORMATION:  
; APPLICANT: PLOWMAN, GREGORY  
; APPLICANT: MARTINEZ, RICARDO  
; APPLICANT: WHYTE, DAVID  
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES  
; FILE REFERENCE: 038602/0329  
; CURRENT APPLICATION NUMBER: US/09/291,417D  
; CURRENT FILING DATE: 1999-04-13  
; PRIOR FILING DATE: 60/081,784  
; PRIOR FILING DATE: 1998-04-14  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 23  
; LENGTH: 786  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-291-417D-23

Query Match 9.0%; Score 251.5; DB 4; Length 786;  
Best Local Similarity 21.6%; Pred. No. 6.1e-09;  
Matches 139; Conservative 118; Mismatches 196; Indels 191; Gaps 29;  
QY 39 PAVAE-----GPGSSQA-PRKEGAQAARTAQSGALRDVSELSRQLE 80  
DB 109 PAVAEQEEBQDHGVRTVNSVGNQSIPEMSISASSQSSVNSLPDVSDDKS-ELD 167  
QY 81 DILSTYCVDNNGQ-----GPGEDG--AQGEPAEPDAEKSRITYVARNGEPEFTPVNGEKE 134  
DB 168 MMEGDHTWMSNSVHLKPEENYREBGP-----RT---RASDPQSPQVSRHKS 215  
QY 135 PSKGDPTTEIROQSDVGDRRHRRPOEKKKAGLGEKIT-----LIMQTLNTLSTPEEKLA 190

Db 216 HYNREHPTATISLV-----TRQVQEHQDSLEBQMSGYKMRQHQKQLATLENK 271  
 Qy 191 ALCKKYA---ELLEHRN--SOKOMKLOKQSOVLQV-----KDHURGEHSK 233  
 Db 272 AEMDEHRLDLKDTORNFNFAAMEKLIKQQAAMEKAQVMSNEEKFFQOHIQAQOKK 331  
 Qy 234 AVLARSKLESCLRELQHRNRSKKEGVOR---AREEBE---KKEVTSHFQV--TLNDI 284  
 Db 332 EL--NSFLESQKREYKLRKEQLKEHLENQSTPKKEKQEWLSKQENIHFQCAEBEANLL 389  
 Qy 285 QLOME-----QHN-----ERNSKLRQENMELARLKKLIQOYELRE 320  
 Db 390 RRQRYLELCRRFRKRRMLGRHNLQDLVREELNKQTKQKLEHAMLROHESMOSELEF 449  
 Qy 321 EHDVKVFKHDLQOOLVDKILQQAQEMKAEERHOREKDFLLKEAVESQRCMLKQOE 380  
 Db 450 RHNLTIQK--MRCELI--RLQHOTELTNQLEYNKRREBLRRKHVMEVROQPKSLKSKE 504  
 Qy 381 THLKOOL-----ALYTEKPEEPONTLSKSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 436  
 Db 505 LQIKKQFQDTCKIOTROYKALRNHLE-----TTPKSEHKAVLKRLK--EQTRKLAILA 557  
 Qy 437 ESSNKALLEMAEKTVR-----DKELE 458  
 Db 558 EYDHSINEMSLTOALRLDEAQAECQVLKMOLOQLELLNAYQSKIKMQAQAQHDREL 617  
 Qy 459 GLOVKI-----ORLEKLCRALQTERND-----LNKRVQDLSAGGGS-----L 496  
 Db 618 ELEQVSLRRALLEQKIEEMALQNERIRSLRQAREIEAFDSSEYELGFSNMVL 677  
 Qy 497 TDSGPRRPEGGAQAPSPRVTEAPCYCAP--STEASGQTGP 538  
 Db 678 SNLSPE-----APSHSYPGASGWSHNPTGSGP 705

## RESULT 13

US-09-688-1888-31  
 ; Sequence 31, Application US/096881888  
 ; Patent No. 6656716  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PLOWMAN, GREGORY  
 ; APPLICANT: MARTINEZ, RICARDO  
 ; APPLICANT: WHYTE, DAVID  
 ; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES  
 ; FILE REFERENCE: 038602/0328  
 ; CURRENT APPLICATION NUMBER: US/09/688,1888  
 ; CURRENT FILING DATE: 2000-10-16  
 ; PRIOR APPLICATION NUMBER: 09/291,417  
 ; PRIOR FILING DATE: 1999-04-14  
 ; PRIOR APPLICATION NUMBER: 60/081,784  
 ; PRIOR FILING DATE: 1998-04-14  
 ; NUMBER OF SEQ ID NOS: 155  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 31  
 ; LENGTH: 1001  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 ; OTHER INFORMATION: murine/human SULU3  
 US-09-688-1888-31

Query Match 9.0%; Score 251.5; DB 4; Length 1001;  
 Best Local Similarity 21.6%; Pred. No. 8.3e-09;  
 Matches 139; Conservative 118; Mismatches 196; Indels 191; Gaps 29;  
 Qy 39 PAVEAB-----GPGSSQA-PRKEGQAQRTAQSALDVSSELSRQLE 80  
 Db 324 PAVEAEEEBEODHGVGRGTGTVNSVGSNQSIPTSMISASQSSSVNSLDPVSDDKS-ELD 382  
 Qy 81 DILSTVCVNNQO---GPGEDG--AQGEPAEPDAEKRTYVARNGEPEPTPVYNGEKE 134  
 Db 383 MVEGDHTVMSNSVTHLKEENYREGDP-----RT---RASDPQSPQVSRHKS 430

Qy 135 PSKGDPTNEITROSDEVDGRDHRPOEKAKKAGLGKEIT-----LLMOTLNTLSTPERKLA 190  
 Db 431 HYNREHPTATISLV-----TRQVQEHQDSLEBQMSGYKMRQHQKQLATLENK 486  
 Qy 191 ALCKKYA---ELLEHRN--SOKOMKLOKQSOVLQV-----KDHURGEHSK 233  
 Db 487 AEMDEHRLDLKDTORNFNFAAMEKLIKQQAAMEKAQVMSNEEKFFQOHIQAQOKK 546  
 Qy 234 AVLARSKLESCLRELQHRNRSKKEGVOR---AREEBE---KKEVTSHFQV--TLNDI 284  
 Db 547 EL--NSFLESQKREYKLRKEQLKEHLENQSTPKKEKQEWLSKQENIHFQCAEBEANLL 604  
 Qy 285 QLOME-----QHN-----ERNSKLRQENMELARLKKLIQOYELRE 320  
 Db 605 RRQRYLELCRRFRKRRMLGRHNLQDLVREELNKQTKQKLEHAMLROHESMOSELEF 664  
 Qy 321 EHDVKVFKHDLQOOLVDKILQQAQEMKAEERHOREKDFLLKEAVESQRCMLKQOE 380  
 Db 665 RHNLTIQK--MRCELI--RLQHOTELTNQLEYNKRREBLRRKHVMEVROQPKSLKSKE 719  
 Qy 381 THLKOOL-----ALYTEKPEEPONTLSKSSEVFTTFKQEMKMTKKIKKLEKETTMYRSRW 436  
 Db 720 LQIKKQFQDTCKIOTROYKALRNHLE-----TTPKSEHKAVLKRLK--EQTRKLAILA 772  
 Qy 437 ESSNKALLEMAEKTVR-----DKELE 458  
 Db 773 EYDHSINEMSLTOALRLDEAQAECQVLKMOLOQLELLNAYQSKIKMQAQAQHDREL 832  
 Qy 459 GLOVKI-----ORLEKLCRALQTERND-----LNKRVQDLSAGGGS-----L 496  
 Db 833 ELEQVSLRRALLEQKIEEMALQNERIRSLRQAREIEAFDSSEYELGFSNMVL 892  
 Qy 497 TDSGPRRPEGGAQAPSPRVTEAPCYCAP--STEASGQTGP 538  
 Db 893 SNLSPE-----APSHSYPGASGWSHNPTGSGP 920

## RESULT 14

US-09-291-417D-31  
 ; Sequence 31, Application US/09291417D  
 ; Patent No. 6680170  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PLOWMAN, GREGORY  
 ; APPLICANT: MARTINEZ, RICARDO  
 ; APPLICANT: WHYTE, DAVID  
 ; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES  
 ; FILE REFERENCE: 038602/0329  
 ; CURRENT APPLICATION NUMBER: US/09/291,417D  
 ; CURRENT FILING DATE: 1999-04-13  
 ; PRIOR APPLICATION NUMBER: 60/081,784  
 ; PRIOR FILING DATE: 1998-04-14  
 ; NUMBER OF SEQ ID NOS: 155  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 31  
 ; LENGTH: 1001  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 ; OTHER INFORMATION: murine/human SULU3  
 US-09-291-417D-31

Query Match 9.0%; Score 251.5; DB 4; Length 1001;  
 Best Local Similarity 21.6%; Pred. No. 8.3e-09;  
 Matches 139; Conservative 118; Mismatches 196; Indels 191; Gaps 29;  
 Qy 39 PAVEAB-----GPGSSQA-PRKEGQAQRTAQSALDVSSELSRQLE 80  
 Db 324 PAVEAEEEBEODHGVGRGTGTVNSVGSNQSIPTSMISASQSSSVNSLDPVSDDKS-ELD 382  
 Qy 81 DILSTVCVNNQO---GPGEDG--AQGEPAEPDAEKRTYVARNGEPEPTPVYNGEKE 134

```

Db 383 MMEGDTWMSNSSVHLKPEENYREEDP-----RT---RASDPQSPPOVSRHKS 430
Qy 135 PSKGDPTNTEIROSDEVGDRHRRPOEKKAKGLGKEIT-----LLMQTLNTLSTPEZKLA 190
Db 431 HYNRREHPTATISLV-----TRQMEHQDESELRBQMSGYKRMRRQKQKQATLENKLGK 486
Qy 191 ALCKKYA-----ELLEHRN--SQOMKLLQKQSQVQ-----KDHRLRGHSHK 233
Db 487 AKMDHRLRLDKLETQRNFAAEMEKLIKKHQAAMEKEAKVMSNEEKKFQOHQAQOKK 546
Qy 234 AVLARSKLESICRELQRNRSIKKEGVQ-----AREEE--KRKEVTSHFQV--TLNDI 284
Db 547 EL--NSFLESQKREYKLAKEBLNENQSTPKKEQEWLSKQENIQHFQAEZEANLL 604
Qy 285 QLOME-----QHN-----ERNSKLRQENMELAEKLLIYOYELRE 320
Db 605 RRQOYLELECRFRERMGLGRHNLQDLVREELNKRQTKDLEHMLLRQHESMQHELP 664
Qy 321 EHIKVFHKLQOOLVDAKLOQAQOEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOE 380
Db 665 RHLNTIQK---MRCELI--RLQHQTETLNQLEYNKRREBELRRKHVMEVROQPKSLAKSE 719
Qy 381 THLKQOL---ALYTEKPEEFONTLSKSSEVFTTFKQEMKEMTKIKKLEKETTMYRSRW 436
Db 720 LQIKKQFQDTCKIQTRYKALRNHLE-----TPKSEHKAVLKLK--EQTRKLAILAQ 774
Qy 437 ESSNKALLEMAEKTVR-----DRELE 458
Db 773 EYDHSINEMLSQALRLDEAQAECQVLKMQLOQLELLNAYQSKIKNQAEQAQHDREL 832
Qy 459 GLOVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGGS 496
Db 833 ELEQRVSRALLFOKIEEMALQNERTERIRSLERQAREIEAFDSMESMLGFSNNVL 892
Qy 497 TDSGPRRPEGPGAQAPSPPRVTPACYPGAP--STEASGQTGP 538
Db 893 SNLSPE-----APSHSYPGASGSHNFTGPGP 920

```

## RESULT 15

```

US-09-060-410-2
; Sequence 2, Application US/09060410
; Patent No. 6165461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Berman, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,410
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

```

```

; TELEPAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1001 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-060-410-2

```

Query Match 9.0%; Score 250; DB 3; Length 1001;

Best local Similarity 21.9%; Pred. No. le-08;  
Matches 142; Conservative 119; Mismatches 211; Indels 176; Gaps 30;

```

Qy 39 PAYEAE-----GPGSSQA--PRKPEGAQARTAQSGALRVSVSELSQLE 80
Db 324 PAVEAQEEBEEQDGGGRTCTVNSVGSNQSIPTMSISASSQSSVNSLPPASDDKS--ELD 382
Qy 81 DIUSTVCVNNQG-----GPGEDGAQGEPAEPEDAESKRTTVARKEPEPTPVVNGEKEPS 136
Db 383 MMEGDTWMSNSSVHLKPEENYQ--EEGDP-----TRASAPQSPPOVSRHSHY 432
Qy 137 KGPNTTEIROSDEVGDRHRRPOEKKAKGLGKEIT-----LLMQTLNTLSTPEEKLAAL 192
Db 433 RNREHPTATISLV-----TRQMEHQDESELRBQMSGYKRMRRQKQKQATLENKLAKE 489
Qy 193 CKYA-----ELLEHRN--SQOMKLLQKQSQVQ-----KDHRLRGHSHKAV 235
Db 489 MDEHRLRLDKLETQRNFAAEMEKLIKKHQAAMEKEAKVMAEKKFQOHQAQOKKEL 548
Qy 236 LARSKLESICRELQRNRSIKKEGVQ-----AREEE--KRKEVTSHFQV--TLNDIOL 286
Db 549 --NSFLESQKREYKLAKEBLNENQSTPKKEQEWLSKQENIQHFQAEZEANLLRR 606
Qy 287 QME-----QHN-----ERNSKLRQENMELAEKLLIYOYELREEH 322
Db 607 QROYLELECRFRERMGLGRHNLQDLVREELNKRQTKDLEHMLLRQHESMQHELSFRH 666
Qy 323 IDKVFHKLQOOLVDAKLOQAQOEMLEKAEERHOREKDFLLKEAVESQRMCELMKQOETH 382
Db 667 LNTIQK---MRCELI--RLQHQTETLNQLEYNKRREBELRRKHVMEVROQPKSLAKSELQ 721
Qy 383 LKQOL---ALYTEKPEEFONTLSKSSEVFTTFKQEMKEMTKIKKLEKETTMYRSRWES 438
Db 722 IKQFQDTCKIQTRYKALRNHLE-----TPKSEHKAVLKLK--EQTRKLAILAQ 774
Qy 439 SNKALLEMAEKTVR-----DKELEGL 460
Db 775 YDHSINEMLSQALRLDEAQAECQVLKMQLOQLELLNAYQSKIKNQAEQAQHDRELREL 834
Qy 461 QVKI-----ORLEKLCRALQTERND-----LNKRVODLSAGGGS 498
Db 835 EQVSLRRALLEOKIEEMALQNERTERIRSLERQAREIEAFDSMESMLGFSNNVLN 894
Qy 499 SGPERRPEG-PGAQAPS-SPRVTEAP--CYPGAPSTEASG---QTPQ 539
Db 895 LSPEAFSHSYPGASGSHNFTGSGPHWGHMGTGTPQAWGHPMGQGP 942

```

Search completed: June 8, 2004, 16:36:45

Job time : 31 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: June 8, 2004, 16:31:45 ; Search time 56 Seconds  
(without alignments)  
2743.052 Million cell updates/sec

Title: US-10-023-529-44  
Perfect score: 2785  
Sequence: 1 MKNQDKNGAKQSNPKSSP.....APSTEASGQTGPQPTPSARA 546

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PTC\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PTCUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	2785	100.0	546	9	US-09-976-740-44
2	2785	100.0	546	12	US-10-671-242-44
3	2785	100.0	546	13	US-10-023-529-44
4	2785	100.0	546	13	US-10-023-523-44
5	2785	100.0	546	15	US-10-616-187-44
6	2693	96.7	530	9	US-09-962-055-8
7	2693	96.7	530	9	US-09-976-740-8
8	2693	96.7	530	12	US-10-671-242-8
9	2693	96.7	530	13	US-10-023-529-8
10	2693	96.7	530	13	US-10-023-523-8
11	2693	96.7	530	15	US-10-616-187-8
12	2552.5	91.7	557	9	US-09-962-055-5
13	2552.5	91.7	557	9	US-09-976-740-5
14	2552.5	91.7	557	12	US-10-671-242-5
15	2552.5	91.7	557	13	US-10-023-529-5

16	2552.5	91.7	557	13	US-10-023-523-5
17	2552.5	91.7	557	15	US-10-616-187-5
18	2416	86.8	510	12	US-10-276-774-2134
19	501	16.9	437	12	US-10-425-114-55188
20	470.5	16.9	436	12	US-10-425-599-207434
21	444.5	16.0	314	12	US-10-425-114-43684
22	375	13.5	346	15	US-10-369-493-5130
23	304.5	10.9	206	12	US-10-425-114-45419
24	263	9.4	2871	14	US-10-146-473-41
25	260	9.3	886	15	US-10-369-493-1016
26	260	9.3	909	9	US-09-925-299-988
27	260	9.3	909	10	US-09-925-299-988
28	257.5	9.2	860	12	US-10-072-012-838
29	257.5	9.2	860	12	US-10-037-417-59
30	257.5	9.2	860	15	US-10-080-334-166
31	254.5	9.1	880	15	US-10-369-493-21643
32	253.5	9.1	1534	14	US-10-342-136-1
33	252.5	9.1	1881	14	US-10-032-585-7646
34	251.5	9.0	786	10	US-09-291-417-31
35	251.5	9.0	1001	10	US-09-291-417-31
36	251.5	9.0	1005	14	US-10-177-293-254
37	251.5	9.0	1005	14	US-10-317-835-18
38	250	9.0	1001	15	US-10-445-735-2
39	247	8.9	85	15	US-10-264-049-3413
40	244.5	8.8	3225	16	US-10-408-765A-254
41	240	8.6	650	15	US-10-104-047-3636
42	240	8.6	1001	14	US-10-128-714-3240
43	240	8.6	1162	12	US-09-894-273-2
44	240	8.6	1162	14	US-10-294-804-2
45	239	8.6	1137	12	US-10-336-472-16

ALIGNMENTS

RESULT 1

US-09-976-740-44  
; Sequence 44, Application US/09976740  
; Publication No. US20020194633A1  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/09/976,740  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 09/616,289  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; PRIOR FILING DATE: 1997-06-03  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 546  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-976-740-44

Query Match 100.0%; Score 2785; DB 9; Length 546;  
Best Local Similarity 100.0%; Pred. No. 1.2e-156;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNQDKNGAKQSNPKSPGQPEAGPQAGPQPSQAAPAVEAGPGSSQAPKPEGAQA 60  
DB 1 MKNQDKNGAKQSNPKSPGQPEAGPQAGPQPSQAAPAVEAGPGSSQAPKPEGAQA 60

Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKQDKNGAAGKSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKEGAQA 60  
 Db 1 MKQDKNGAAGKSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKEGAQA 60

Qy 61 RTAQSALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGPAPEDAEKSTYVARN 120  
 Db 61 RTAQSALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGPAPEDAEKSTYVARN 120

Qy 121 GEPTPTVNVGKESKSPGDPNTEIRQSDVGVDRDRHRRPOEKKAKGLGKEITLLMOTLN 180  
 Db 121 GEPTPTVNVGKESKSPGDPNTEIRQSDVGVDRDRHRRPOEKKAKGLGKEITLLMOTLN 180

Qy 181 TLSTPEEKLAALCKKYAELLBEHNSQOMKLLQKKOSQOLVQEKDHLRGHSHKAVLARSK 240  
 Db 181 TLSTPEEKLAALCKKYAELLBEHNSQOMKLLQKKOSQOLVQEKDHLRGHSHKAVLARSK 240

Qy 241 LESICRELQNRNSLKEGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300  
 Db 241 LESICRELQNRNSLKEGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300

Qy 301 ENMELAEKRLKLIQYELREBEHIDKVFVKHDLQOOLVDKLOQAQEMLKEAERHOREKD 360  
 Db 301 ENMELAEKRLKLIQYELREBEHIDKVFVKHDLQOOLVDKLOQAQEMLKEAERHOREKD 360

Qy 361 FLKKEAVESQRMCELMKQOETHLKOQALYTERKEFEFQNTLSKSEVFTTFKQEMEKMTK 420  
 Db 361 FLKKEAVESQRMCELMKQOETHLKOQALYTERKEFEFQNTLSKSEVFTTFKQEMEKMTK 420

Qy 421 KIKKLEKTTMYRSRVESNNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERND 480  
 Db 421 KIKKLEKTTMYRSRVESNNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERND 480

Qy 481 LNKRVQDLSAGGQSLTDSGPERRPGPGAQAPSPRVTEAPCYPGAPSTEASGQTGPQE 540  
 Db 481 LNKRVQDLSAGGQSLTDSGPERRPGPGAQAPSPRVTEAPCYPGAPSTEASGQTGPQE 540

Qy 541 PTSARA 546  
 Db 541 PTSARA 546

RESULT 3  
 US-10-023-529-44  
 ; Sequence 44, Application US/10023529  
 ; Publication No. US20020129388A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lees, Ann M.  
 ; APPLICANT: Lees, Robert S.  
 ; APPLICANT: Law, Simon W.  
 ; APPLICANT: Arjona, Anibal A.  
 ; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
 ; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
 ; FILE REFERENCE: 10797-004001  
 ; CURRENT APPLICATION NUMBER: US/10/023,529  
 ; PRIOR FILING DATE: 2001-12-17  
 ; PRIOR APPLICATION NUMBER: 09/616,289  
 ; PRIOR FILING DATE: 2000-07-14  
 ; PRIOR APPLICATION NUMBER: US 09/517,849  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: US 08/979,608  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 60/031,930  
 ; PRIOR FILING DATE: 1996-11-27  
 ; PRIOR APPLICATION NUMBER: US 60/048,547  
 ; PRIOR FILING DATE: 1997-06-03  
 ; NUMBER OF SEQ ID NOS: 53  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 44  
 ; LENGTH: 546  
 ; TYPE: PRT

Qy 61 RTAQSALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGPAPEDAEKSTYVARN 120  
 Db 61 RTAQSALRDVSEELSRQLEDILSTYCVNNQGGPGEDGAQGPAPEDAEKSTYVARN 120

Qy 121 GEPTPTVNVGKESKSPGDPNTEIRQSDVGVDRDRHRRPOEKKAKGLGKEITLLMOTLN 180  
 Db 121 GEPTPTVNVGKESKSPGDPNTEIRQSDVGVDRDRHRRPOEKKAKGLGKEITLLMOTLN 180

Qy 181 TLSTPEEKLAALCKKYAELLBEHNSQOMKLLQKKOSQOLVQEKDHLRGHSHKAVLARSK 240  
 Db 181 TLSTPEEKLAALCKKYAELLBEHNSQOMKLLQKKOSQOLVQEKDHLRGHSHKAVLARSK 240

Qy 241 LESICRELQNRNSLKEGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300  
 Db 241 LESICRELQNRNSLKEGVQARAREEERKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300

Qy 301 ENMELAEKRLKLIQYELREBEHIDKVFVKHDLQOOLVDKLOQAQEMLKEAERHOREKD 360  
 Db 301 ENMELAEKRLKLIQYELREBEHIDKVFVKHDLQOOLVDKLOQAQEMLKEAERHOREKD 360

Qy 361 FLKKEAVESQRMCELMKQOETHLKOQALYTERKEFEFQNTLSKSEVFTTFKQEMEKMTK 420  
 Db 361 FLKKEAVESQRMCELMKQOETHLKOQALYTERKEFEFQNTLSKSEVFTTFKQEMEKMTK 420

Qy 421 KIKKLEKTTMYRSRVESNNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERND 480  
 Db 421 KIKKLEKTTMYRSRVESNNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQTERND 480

Qy 481 LNKRVQDLSAGGQSLTDSGPERRPGPGAQAPSPRVTEAPCYPGAPSTEASGQTGPQE 540  
 Db 481 LNKRVQDLSAGGQSLTDSGPERRPGPGAQAPSPRVTEAPCYPGAPSTEASGQTGPQE 540

Qy 541 PTSARA 546  
 Db 541 PTSARA 546

RESULT 2  
 US-10-671-242-44  
 ; Sequence 44, Application US/10671242  
 ; Publication No. US20040040049A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lees, Ann M.  
 ; APPLICANT: Lees, Robert S.  
 ; APPLICANT: Law, Simon W.  
 ; APPLICANT: Arjona, Anibal A.  
 ; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
 ; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
 ; FILE REFERENCE: 10797-004001  
 ; CURRENT APPLICATION NUMBER: US/10/671,242  
 ; PRIOR FILING DATE: 2003-03-24  
 ; PRIOR APPLICATION NUMBER: US/09/616,289  
 ; PRIOR FILING DATE: 2000-07-14  
 ; PRIOR APPLICATION NUMBER: US 09/517,849  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: US 08/979,608  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 60/031,930  
 ; PRIOR FILING DATE: 1996-11-27  
 ; PRIOR APPLICATION NUMBER: US 60/048,547  
 ; PRIOR FILING DATE: 1997-06-03  
 ; NUMBER OF SEQ ID NOS: 53  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 44  
 ; LENGTH: 546  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-671-242-44

Query Match 100.0%; Score 2785; DB 12; Length 546;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-156;

; ORGANISM: Homo sapiens  
US-10-023-529-44  
Query Match 100.0%; Score 2785; DB 13; Length 546;  
Best Local Similarity 100.0%; Pred. No. 1.2e-156;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; SEQ ID NO 44  
; LENGTH: 546  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-523-44  
Query Match 100.0%; Score 2785; DB 13; Length 546;  
Best Local Similarity 100.0%; Pred. No. 1.2e-156;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
; SEQ ID NO 44  
; LENGTH: 546  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-023-529-44

1 MNQDKKNGAAKQSNPKSPGQPEAGPGEQAQRPSPQAAPAVEAEGPGSSQAAPKPEGAQA 60  
Db 1 MNQDKKNGAAKQSNPKSPGQPEAGPGEQAQRPSPQAAPAVEAEGPGSSQAAPKPEGAQA 60  
61 RTAQSGALRDVSEELSRQLEDILSTVCVNNQGGPGEDGAQGPAPEDAERKSRVYARN 120  
Db 61 RTAQSGALRDVSEELSRQLEDILSTVCVNNQGGPGEDGAQGPAPEDAERKSRVYARN 120  
121 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVGVDRDHRPQEKKAKGLGKEITLLMQTLN 180  
Db 121 GEPEPTPVVNGEKEPSKGDPTNTEIRQSDVGVDRDHRPQEKKAKGLGKEITLLMQTLN 180  
181 TLSTPEKLAALCKYAEALLLEHRNSQKMLLOKQSQQLVQEKDHLRGHSHKAVLARSK 240  
Db 181 TLSTPEKLAALCKYAEALLLEHRNSQKMLLOKQSQQLVQEKDHLRGHSHKAVLARSK 240  
241 LESLCRELQHRNSLKEEGVQARAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300  
Db 241 LESLCRELQHRNSLKEEGVQARAREEKEKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300  
301 ENNELAERLKKLIQEVYELREHIDKVFHKDLQOQLVDAKLQQAQEMLKEAERHOREKD 360  
Db 301 ENNELAERLKKLIQEVYELREHIDKVFHKDLQOQLVDAKLQQAQEMLKEAERHOREKD 360  
361 FLKKEAVESQRMCELMKQOETHLKOOLALYTERFEFQNTLSKSSSEVFTTFKQEMEKMTK 420  
Db 361 FLKKEAVESQRMCELMKQOETHLKOOLALYTERFEFQNTLSKSSSEVFTTFKQEMEKMTK 420  
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKSLQGLQVKIQRLEKLCRALQTERND 480  
Db 421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVDKSLQGLQVKIQRLEKLCRALQTERND 480  
481 LNKRVQDLSAGGQSLTDSGPERPPEGQAQAPSPRVTEAPCYGAPSTASGQTGPQE 540  
Db 481 LNKRVQDLSAGGQSLTDSGPERPPEGQAQAPSPRVTEAPCYGAPSTASGQTGPQE 540  
541 PTSARA 546  
Db 541 PTSARA 546  
RESULT 4  
US-10-023-523-44  
; Sequence 44, Application US/10023523  
; Publication No. US20020152485A1  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/10/023,523  
; CURRENT FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US/09/616,289  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 09/517,849  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; PRIOR FILING DATE: 1997-06-03

; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 60/031,930  
 ; PRIOR FILING DATE: 1996-11-27  
 ; PRIOR APPLICATION NUMBER: US 60/048,547  
 ; PRIOR FILING DATE: 1997-06-03  
 ; NUMBER OF SEQ ID NOS: 53  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 44  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-616-187-44

Query Match 100.0%; Score 2785; DB 15; Length 546;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-156;  
 Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKNQDKXGAKQNPSSPQGPAGPEGAQERPSQAAPVAEAGPGSSQAPRKEGAQA	60
Db	1	MKNQDKXGAKQNPSSPQGPAGPEGAQERPSQAAPVAEAGPGSSQAPRKEGAQA	60
Qy	61	RTAQSGLRDVSEBELSRLQLEILSTYCVNNQGGPGEDGAQEPAPEDAKSRTYVARN	120
Db	61	RTAQSGLRDVSEBELSRLQLEILSTYCVNNQGGPGEDGAQEPAPEDAKSRTYVARN	120
Qy	121	GEPEPTVVGKEKPSKDPNTEIRQSDVGVDRHRRPQKKAKGLGKEITLLMOTLN	180
Db	121	GEPEPTVVGKEKPSKDPNTEIRQSDVGVDRHRRPQKKAKGLGKEITLLMOTLN	180
Qy	181	TLSTPEEKLAALCKKYAELLBEHRNSQKMLQKQSQVQKDHRLRGEHSKAVLARS	240
Db	181	TLSTPEEKLAALCKKYAELLBEHRNSQKMLQKQSQVQKDHRLRGEHSKAVLARS	240
Qy	241	LESICRELRNRSIKKEGVORAREEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQ	300
Db	241	LESICRELRNRSIKKEGVORAREEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQ	300
Qy	301	ENMELAEKRLKLEQYELREEHIDKVFHKDLQQLVDKLLQQAQEMLEAEERHOREKD	360
Db	301	ENMELAEKRLKLEQYELREEHIDKVFHKDLQQLVDKLLQQAQEMLEAEERHOREKD	360
Qy	361	FLKKEVTSORVCELMKQOETHLQKQALYTEKFEFQNTLSKSEVTTTQKQEMKTK	420
Db	361	FLKKEVTSORVCELMKQOETHLQKQALYTEKFEFQNTLSKSEVTTTQKQEMKTK	420
Qy	421	KIKLEKETTYRSRWESNKALEMAEKTVRDKEGLQVQKIQLEKLCRALQTERND	480
Db	421	KIKLEKETTYRSRWESNKALEMAEKTVRDKEGLQVQKIQLEKLCRALQTERND	480
Qy	481	LNRVQDLSAGQGSITDGGPRRPEGGAQAPSSPRVTEAPCTPGAPSTASGQTGPQE	540
Db	481	LNRVQDLSAGQGSITDGGPRRPEGGAQAPSSPRVTEAPCTPGAPSTASGQTGPQE	540
Qy	541	PTSARA 546	
Db	541	PTSARA 546	

RESULT 6  
 US-09-962-055-8  
 ; Sequence 8, Application US/09962055  
 ; Patent No. US20020052033A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lees, Ann M.  
 ; Lees, Robert S.  
 ; Law, Simon W.  
 ; Arjona, Anibal A.  
 ; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
 ; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
 ; TREATING ATHEROSCLEROSIS  
 ; NUMBER OF SEQUENCES: 42  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION NUMBER: US/09/962,055  
 FILING DATE: 24-Sep-2001  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/979,608  
 FILING DATE: 26-NOV-1997  
 APPLICATION NUMBER: US 60/031,930  
 FILING DATE: 27-NOV-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Myers, Louis  
 REGISTRATION NUMBER: 35,965  
 REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 530 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
 US-09-962-055-8

Query Match 96.7%; Score 2693; DB 9; Length 530;  
 Best Local Similarity 99.8%; Pred. No. 3.1e-151;  
 Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	17	KSSPGPPEAGPEGAQERPSQAAPVAEAGPGSSQAPRKEGAQA	76
Db	1	KSSPGPPEAGPEGAQERPSQAAPVAEAGPGSSQAPRKEGAQA	60
Qy	77	RLQEDILSTYCVNNQGGPGEDGAQEPAPEDAKSRTYVARNQGPPEPTPVVNGKEPS	136
Db	61	RLQEDILSTYCVNNQGGPGEDGAQEPAPEDAKSRTYVARNQGPPEPTPVVNGKEPS	120
Qy	137	KGDPNTEIRQSDVGVDRHRRPQKKAKGLGKEITLLMOTLN	196
Db	121	KGDPNTEIRQSDVGVDRHRRPQKKAKGLGKEITLLMOTLN	180
Qy	197	AELEEHNRNSOKMLQKQSQVQKDHRLRGEHSKAVLARS	256
Db	181	AELEEHNRNSOKMLQKQSQVQKDHRLRGEHSKAVLARS	240
Qy	257	EEGVQARAREEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQ	316
Db	241	EEGVQARAREEERKEVTSHPQVTLNDIQLQMEQHNRNSKLQ	300
Qy	317	ELREEHIDKVFHKDLQQLVDKLLQQAQEMLEAEERHOREKD	376
Db	301	ELREEHIDKVFHKDLQQLVDKLLQQAQEMLEAEERHOREKD	360
Qy	377	KQETHLQKQALYTEKFEFQNTLSKSEVTTTQKQEMKTK	436
Db	361	KQETHLQKQALYTEKFEFQNTLSKSEVTTTQKQEMKTK	420
Qy	437	ESSNKALLEMAEKTVRDKEGLQVQKIQLEKLCRALQTERND	496
Db	421	ESSNKALLEMAEKTVRDKEGLQVQKIQLEKLCRALQTERND	480
Qy	497	TDGSPERRRPEGGAQAPSSPRVTEAPCTPGAPSTASGQTGPQE	546
Db	481	TDGSPERRRPEGGAQAPSSPRVTEAPCTPGAPSTASGQTGPQE	530

## RESULT 7

US-09-976-740-8  
; Sequence 8, Application US/09976740  
; Publication No. US20020194633A1  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/09/976,740  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 09/616,289  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; PRIOR FILING DATE: 1997-06-03  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 530  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-976-740-8

Query Match 96.7%; Score 2693; DB 9; Length 530;  
Best Local Similarity 99.8%; Pred. No. 3.1e-151;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAAPRKEGAQAARTAQSGALRDVSEELS 76  
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAAPRKEGAQAARTAQSGALRDVSEELS 60  
  
Qy 77 ROLEDILSTYCVDNNGGPGEDGAGPEPAEDAEKSRITYVARNGEPEPTPVVYGEKPS 136  
Db 61 ROLEDILSTYCVDNNGGPGEDGAGPEPAEDAEKSRITYVARNGEPEPTPVVYGEKPS 120  
  
Qy 137 KGDPTTEIRQSDVGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 196  
Db 121 KGDPTTEIRQSDVGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180  
  
Qy 197 AELEEHRSNOKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESCLRELQHRNRSK 256  
Db 181 AELEEHRSNOKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESCLRELQHRNRSK 240  
  
Qy 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEERLKLIEQY 316  
Db 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEERLKLIEQY 300  
  
Qy 317 ELREEHIDKVFKHDLQQLVDKLLQQAQEMLKEAERHOREKDPLLKEAVESQRMCELM 376  
Db 301 ELREEHIDKVFKHDLQQLVDKLLQQAQEMLKEAERHOREKDPLLKEAVESQRMCELM 360  
  
Qy 377 KQETHLKKQALALYTEKPEFQNTLSKSSVFTTFKQEMEKMTKKIKLEKETTYRSRW 436  
Db 361 KQETHLKKQALALYTEKPEFQNTLSKSSVFTTFKQEMEKMTKKIKLEKETTYRSRW 420  
  
Qy 437 ESSNKALLEMAEKTVRDKELEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSL 496  
Db 421 ESSNKALLEMAEKTVRDKELEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSL 480  
  
Qy 497 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 546  
Db 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530

## RESULT 8

US-10-671-242-8  
; Sequence 8, Application US/10671242  
; Publication No. US20040040049A1  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/10/671,242  
; PRIOR FILING DATE: 2003-09-24  
; PRIOR APPLICATION NUMBER: US/09/616,289  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 09/517,849  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; PRIOR FILING DATE: 1997-06-03  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 530  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-671-242-8

Query Match 96.7%; Score 2693; DB 12; Length 530;  
Best Local Similarity 99.8%; Pred. No. 3.1e-151;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 17 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAAPRKEGAQAARTAQSGALRDVSEELS 76  
Db 1 KSSPGQPEAGPEGAQERPSQAAPAVEAEGPSSQAAPRKEGAQAARTAQSGALRDVSEELS 60  
  
Qy 77 ROLEDILSTYCVDNNGGPGEDGAGPEPAEDAEKSRITYVARNGEPEPTPVVYGEKPS 136  
Db 61 ROLEDILSTYCVDNNGGPGEDGAGPEPAEDAEKSRITYVARNGEPEPTPVVYGEKPS 120  
  
Qy 137 KGDPTTEIRQSDVGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 196  
Db 121 KGDPTTEIRQSDVGRDHRPQEKKAKGLGKEITLLMOTLNTLSTPEEKLAALCKKY 180  
  
Qy 197 AELEEHRSNOKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESCLRELQHRNRSK 256  
Db 181 AELEEHRSNOKMKLLQKQSQQLVQEKDHLRGEHSKAVLARSKLESCLRELQHRNRSK 240  
  
Qy 257 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEERLKLIEQY 316  
Db 241 EGVQARAEERKEKVTSHFQVTLNDIQLQMEQHNRNSKLRQENMELAEERLKLIEQY 300  
  
Qy 317 ELREEHIDKVFKHDLQQLVDKLLQQAQEMLKEAERHOREKDPLLKEAVESQRMCELM 376  
Db 301 ELREEHIDKVFKHDLQQLVDKLLQQAQEMLKEAERHOREKDPLLKEAVESQRMCELM 360  
  
Qy 377 KQETHLKKQALALYTEKPEFQNTLSKSSVFTTFKQEMEKMTKKIKLEKETTYRSRW 436  
Db 361 KQETHLKKQALALYTEKPEFQNTLSKSSVFTTFKQEMEKMTKKIKLEKETTYRSRW 420  
  
Qy 437 ESSNKALLEMAEKTVRDKELEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSL 496  
Db 421 ESSNKALLEMAEKTVRDKELEGLQVKIQRLKLCRALQTERNDLNKRVQDLSAGGQSL 480  
  
Qy 497 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 546  
Db 481 TDSGPRRPEGGAQAPSSPRVTEAPCYPGAPSTEASGQTGPQEPPTSARA 530



Db 481 TDSGPERPEGGAQAPSSPRVTEAPCYGAPSTASQOTGPQEPSTARA 530  
|||||

## RESULT 11

US-10-616-187-8

; Sequence 8, Application US/10616187

; Publication No. US20040013668A1

; GENERAL INFORMATION:

; APPLICANT: Lees, Ann M.

; APPLICANT: Lees, Robert S.

; APPLICANT: Law, Simon W.

; APPLICANT: Arjona, Anibal A.

; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING

; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING

; TITLE OF INVENTION: ATHEROSCLEROSIS

; FILE REFERENCE: 10797-004001

; CURRENT APPLICATION NUMBER: US/10/616,187

; CURRENT FILING DATE: 2003-07-09

; PRIOR APPLICATION NUMBER: US/09/616,289

; PRIOR FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: US 09/517,849

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 08/979,608

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: US 60/031,930

; PRIOR FILING DATE: 1996-11-27

; PRIOR APPLICATION NUMBER: US 60/049,547

; PRIOR FILING DATE: 1997-06-03

; NUMBER OF SEQ ID NOS: 53

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 530

; TYPES: PRT

; ORGANISM: Homo sapiens

; US-10-616-187-8

Query Match 96.7%; Score 2693; DB 15; Length 530;  
Best Local Similarity 99.8%; Pred. No. 3.1e-151;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 17 KSSGQPEAGPEGGAQAPSSPRVTEAPCYGAPSTASQOTGPQEPSTARA 76  
Db 1 KSSGQPEAGPEGGAQAPSSPRVTEAPCYGAPSTASQOTGPQEPSTARA 60

Qy 77 RQEDILSTYCVDDNQGPGEDGAGGEPAPEDAEKSTYVARNGEPEPTPVVYGEKPS 136  
Db 61 RQEDILSTYCVDDNQGPGEDGAGGEPAPEDAEKSTYVARNGEPEPTPVVYGEKPS 120

Qy 137 KQDPNTEIROSDEVDGDRHRRPQKKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 196  
Db 121 KQDPNTEIROSDEVDGDRHRRPQKKKAKGLGKEITLLMQTLNTLSTPEKLAALCKKY 180

Qy 197 AELLREHENSQOMKLLQKQSQOLVQEKDHLRGHESKAVLARSKLESICRBLQRHNSLK 256  
Db 181 AELLREHENSQOMKLLQKQSQOLVQEKDHLRGHESKAVLARSKLESICRBLQRHNSLK 240

Qy 257 EGVQARPEESKREKVTSHFQVTLNDIQLQMEQHNENSKLRQNMELAKLLIQY 316  
Db 241 EGVQARPEESKREKVTSHFQVTLNDIQLQMEQHNENSKLRQNMELAKLLIQY 300

Qy 317 ELREBHIDKVFQKDLQOOLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 376  
Db 301 ELREBHIDKVFQKDLQOOLVDAKLQQAQEMLKEAERHOREKDFLLKEAVESQRMCELM 360

Qy 377 KQOETHLKOQALYTEKPEFQNTLSKSEVPTTFQEMKMTKKIKLEKETMYRSRW 436  
Db 361 KQOETHLKOQALYTEKPEFQNTLSKSEVPTTFQEMKMTKKIKLEKETMYRSRW 420

Qy 437 ESSNKAILEWAEKTVRDELSGLQIKOLEKLCALQTERNDLNKRQDLSAGQGSLL 496  
Db 421 ESSNKAILEWAEKTVRDELSGLQIKOLEKLCALQTERNDLNKRQDLSAGQGSLL 480

Qy 497 TDSGPERPEGGAQAPSSPRVTEAPCYGAPSTASQOTGPQEPSTARA 546  
Db 481 TDSGPERPEGGAQAPSSPRVTEAPCYGAPSTASQOTGPQEPSTARA 530  
|||||

## RESULT 12

US-09-962-055-5

; Sequence 5, Application US/09962055

; Patent No. US20020052033A1

; GENERAL INFORMATION:

; APPLICANT: Lees, Ann M.

; APPLICANT: Lees, Robert S.

; APPLICANT: Law, Simon W.

; APPLICANT: Arjona, Anibal A.

; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN  
; BINDING PROTEINS AND THEIR USES IN DIAGNOSING AND  
; TREATING ATHEROSCLEROSIS

; NUMBER OF SEQUENCES: 42

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish &amp; Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/962,055

; FILING DATE: 24-Sep-2001

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/979,608

; FILING DATE: 26-NOV-1997

; APPLICATION NUMBER: US 60/031,930

; FILING DATE: 27-NOV-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Myers, Louis

; REGISTRATION NUMBER: 35,965

; REFERENCE/DOCKET NUMBER: 10797-002001 (formerly 3983/59818)

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 557 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-962-055-5

Query Match 91.7%; Score 2552.5; DB 9; Length 557;  
Best Local Similarity 91.0%; Pred. No. 6.7e-143;  
Matches 508; Conservative 12; Mismatches 25; Indels 13; Gaps 3;

Qy 1 MNQDKNKGAQKQSNPKSSPCQPEAGPEGGAQAPSSQAPAEAGPGSSQAPRKEPGAQA 60  
Db 1 MNQDKNKGAQKQSNPKSSPCQPEAGPEGGAQAPSSQAPAEAGPGSSQAPRKEPGAQA 59

Qy 61 RTAQSGALRDVSEHLSRQLIEDILSTYCVDDNQGPGEDGAGGEPAPEDAEKSTYVARN 120  
Db 60 RTAQSGALRDVSEHLSRQLIEDILSTYCVDDNQGPGEDGAGGEPAPEDAEKSTYVARN 119

Qy 121 GEPEBP-TPVNGKEPKSGDPTNTEIIRQSDVEGDRDHRPQKKKAKGLGKEITLLMQTL 179  
Db 120 GEPEBP-TPVNGKEPKSGDPTNTEIIRQSDVEGDRDHRPQKKKAKGLGKEITLLMQTL 179

Qy 180 NTLSTPEBKLAALCKKYAELEEHNSOKMKLLQKQSQOLVQEKDHLRGHESKAVLARS 239

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:31:00 ; Search time 20 Seconds  
(without alignments)  
2626.029 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 2785

Sequence: 1 MKWDKXNGAKQSNPKSSP.....APSTEASGQTGPQPTTSARA 546

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1244	44.7	676	JC7222	77K muscle-derived
2	375	13.5	346	F87844	protein T22C1.6 [i
3	372.5	13.4	335	T25110	hypothetical prote
4	281	10.1	992	T46337	hypothetical prote
5	278	10.0	892	T50985	related to transcr
6	263	9.4	2677	A38194	desmoplakin I - hu
7	261.5	9.4	1407	T34021	trichohyalin - rab
8	260	9.3	886	H69378	conserved hypothet
9	258	9.3	1206	T34021	protein kinase SK2
10	257	9.2	955	S24348	myosin heavy chain
11	254.5	9.1	880	F75103	conserved hypothet
12	254	9.1	936	S39083	myosin heavy chain
13	254	9.1	1938	JX0178	myosin heavy chain
14	253.5	9.1	1534	T456734	ribosome receptor,
15	253.5	9.1	1549	A40691	trichohyalin - she
16	250.5	9.0	1233	T14157	serine/threonine p
17	250.5	9.0	2139	T18296	myosin heavy chain
18	250	9.0	1001	T11365	serine/threonine p
19	249	8.9	1392	A45336	microtubule-vesicl
20	249	8.9	1938	A52993	skeletal myosin he
21	247	8.9	1231	T18532	serine/threonine pr
22	246	8.8	1427	S22695	restin - human
23	246	8.8	1937	T38055	myosin heavy chain
24	244.5	8.8	3225	I52300	giantin - human
25	244.5	8.8	3259	A56539	giantin - human
26	243	8.7	1940	A25320	myosin heavy chain
27	242.5	8.7	1898	A45973	trichohyalin - hum
28	242.5	8.7	4687	A39638	plectin - rat
29	240.5	8.6	1934	I48153	myosin heavy chain

## ALIGNMENTS

### RESULT 1

JC7222

77K muscle-derived protein - chicken

C:Species: Gallus gallus (chicken)

C>Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000

C:Accession: JC7222

R:Uyeda, A.; Fukui, I.; Fujimori, K.; Kiyosue, K.; Nishimune, H.; Kasai, M.; Taguchi, T

Biochem. Biophys. Res. Commun. 269, 564-569, 2000

A:Title: MDP77: A novel neurite-outgrowth-promoting protein predominantly expressed in

A:Reference number: JC7222; MUID:20175243; PMID:10708594

A:Accession: JC7222

A:Molecule type: mRNA

A:Residues: 1-676 <UVB>

A:Cross-references: GB:D89999; NID:g7619883; PIDN:BAA94755.1; PID:g7619884

A:Experimental source: crus muscle

C:Comment: This protein, a glycoprotein and a neurite-outgrowth-promoting protein, is

C:Keywords: coiled coil; Glycoprotein; leucine zipper; muscle

Query Match 44.7%; Score 1244; DB 2; Length 676;

Best Local Similarity 54.3%; Pred. No. 1.9e-44;

Matches 260; Conservative 81; Mismatches 104; Indels 34; Gaps 8;

Qy	12	KQSNPKSPQPGAGPEGAERPSQAAPAVEAGPGSSQAPRKEGAQTAQSGALRDV	71
Db	9	KQQQVTSPTQDQGG-----QSKAEFVPSQPLSP-TNQTSAQPEMATC-----DI	53
Qy	72	SEELSRQLEDILSTYCVDDNQGPGEDGA---QGEPAEPEDAEXSRTYVARNGEPPTP	127
Db	54	SEELNRQLEDILKTY-----GSAASLVEKEGTTAETDKPEKEDVGSMDAECED--	102
Qy	128	VVNGEKPSKQDPNTEIRQSDEVDGDRHRRPOBKTKAKGLGKHTLLMQTLNLTSTPEE	187
Db	103	-VNEESEKDKPAPG-DASRAKEPSASKEQK--LEKTLKGLGKEATLLMQSLNKLTPPEE	158
Qy	188	KLAALCKKYAEELLEHEHNSOKMKLLOKQSOLVORNDHLRGHSKAVLARSKLESICRE	247
Db	159	KLDLLFKYAEELLEHEHRAEQKLYLQKRAQITKEKDLQSEHSRAILARSKLESICRE	218
Qy	248	LQRNRLSKEGVQVQARAEERKEKVTSHFQVTLNDIQLQVQCHNERNKSLQRQNMELAE	307
Db	219	LQRNKLTKBETIQARAEERKEKKEITNHFQGLTSLRQAIIEQOQSERNMKLCQENTELAE	278
Qy	308	RLKKLIEQYELREEHIDKVPKHDLQOLVDKILQQAQEMLKAEERHOREKDFLLKEAV	367
Db	279	KLKSIIQYELREEHLDKIPKRELQOKLVDKILQSEKMKAEERHOREKDFLLQQA	338
Qy	368	ESQRMCELMATQEQETHLAKQALATYKPEEFONTLSKSEVFTTFKQEMEKMTKKILEK	427
Db	339	EWKLQAKMLKEQETVLAQITLYSERFEERFKTLTKSNEVFATPKQEMEKMTKKMKLEK	398
Qy	428	ETTYRWRSNSSKALLEMAEKTVRQKELGLOVKIQRLKLCALQTEENDLNKRVQ	486

364K Golgi complex  
slow myosin heavy  
centromere protein  
plectin [imported]  
microtubule bindin  
plectin - human  
synaptonemal compl  
hypothetical prote  
conserved hypothet  
early endosome ant  
myosin alpha heavy  
myosin beta heavy  
myosin alpha heavy  
alpha cardiac myos  
trichohyalin like  
embryonic muscle m



```
Db 399 DTATWKSRFENCNRRALLDMIEBKAMRTKEYCFVLKIQRLNLCRALQBERNELYRKIK 457
RESULT 2
F87844
protein T22C1.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: F87844
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: F87844
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>
A:Cross-references: GB:chr_1; PIDN:CAA99923.1; PID:G9880029; GSPDB:GN00019; CESP:T22C1.6
C:Genetics:
A:Gene: T22C1.6
A:Map position: 1

Query Match 13.5%; Score 375; DB 2; Length 346;
Best Local Similarity 30.8%; Pred. No. 5e-09;
Matches 105; Conservative 69; Mismatches 123; Indels 44; Gaps 9;

QY 166 KGLGKEITLMLQ-----TLNTL-STPE-EKLAALCKKYAELEHNRNSQOMKLLQKQ 217
Db 3 KNFGKLSIPFNEGDEAALLKSLGVPDAEKVQKLIKLAEE--SEKQNAELKIKVLDYDK 60

QY 218 SOLVQEKHLRHEHSAVL-----ARSKLESICRLQHRNRLKEEGVQVQARAEKKEK 272
Db 61 VVKVQDLTEKLRNQIILLTTEAKSLKLEELCRGLQKANHOTREACAKMKKLEVERGL 120

QY 273 VTSHFQVTLNDIQLOMEQHNERNKLRQENMELARLKKLIQYB-----LRSE 321
Db 121 AVSQLVKTLKLEKTWAGERSKSDSLAEDNKKLSKFSFGHYEKMKVTDQIQKEK 180

QY 322 HDVVKFKHKLQQLVDAKLOQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOET 381
Db 181 YWEEYGTNDLILKLTAKLESASIQVKS-----GMEKDELAKIMLEETARVGGALKTEK 236

QY 382 HLKQQLALYTERKEFFQNTLSKSSVFTTFKQEMKMTKIKKLEKTTMYRSWESSNK 441
Db 237 ALREQVQVSAKYSILTSCLSKSNEAFDKFDKDEISRNVNKKQVKEGLSYKKKSDANK 296

QY 442 ALL-----EMAEKTVRDKLEGLQVQIKQLEKLCRALQ 475
Db 297 KVLVLTMTNQEYAEKIATSDK-----KIQMLENLCRALR 330

RESULT 3
T25110
hypotheical protein T22C1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25110
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19982
A:Accession: T25110
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-335 <WIL>
A:Cross-references: EMBL:Z75550; PIDN:CAA99923.2; GSPDB:GN00019; CESP:T22C1.6
A:Experimental source: clone T22C1
C:Genetics:
A:Gene: CESP:T22C1.6
A:Map position: 1
A:Introns: 6/1; 46/3; 66/3; 91/3; 285/2
```

```
Query Match 13.4%; Score 372.5; DB 2; Length 335;
Best Local Similarity 30.8%; Pred. No. 6.1e-09;
Matches 103; Conservative 70; Mismatches 120; Indels 41; Gaps 9;

QY 168 LGKEI--TLLMQTLTSTPE-EKLAALCKKYAELEHNRNSQOMKLLQKQSQOLVQEK 224
Db 1 MGNFDEAALLKSLR--GVDAEKVQKLIKLAEE--SEKQNAELKIKVLDYDKVVKVQDL 56

QY 225 DHLRGEHSKAVL-----ARSKLESICRLQHRNRLKEEGVQVQARAEKKEKVTSHQV 279
Db 57 TEKLRNNTQLLTTEAKSLKLEELCRGLQKANHOTREACAKMKKLEVERGLAVEQLKV 116

QY 280 TLNDIQLOMEQHNERNKLRQENMELARLKKLIQYB-----LRSEHIDKVPK 328
Db 117 TLKDIETKTAAGRSKSDSLAEDNKKLSKFSFGHYEKMKVTDQIQKEKKEKWEYVGK 176

QY 329 HKLQQLVDAKLOQAQEMLKAEERHOREKDFLLKEAVESQRMCELMKQOETHLQOLA 388
Db 177 TKOLEIKLLTAKLESASIQVKS-----GMEKDELAKIMLEETARVGGALKTEKALRQVQ 232

QY 389 LYTEKFEFQNTLSKSSVFTTFKQEMKMTKIKKLEKTTMYRSWESSNKALL----- 444
Db 233 EYSAKYSILTSCLSKSNEAFDKFDKDEISRNVNKKQVKEGLSYKKKSDANKKVLVLT 292

QY 445 ---EMAEKTVRDKLEGLQVQIKQLEKLCRALQ 475
Db 293 TNQEYAEKIATSDK-----KIQMLENLCRALR 319

RESULT 4
T46337
hypotheical protein DKF2p3402413.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46337
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23037
A:Accession: T46337
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-992 <AAA>
A:Cross-references: EMBL:AL137265
A:Experimental source: adult testis; clone DKF2p3402413
C:Genetics:
A:Note: DKF2p3402413.1

Query Match 10.1%; Score 281; DB 2; Length 992;
Best Local Similarity 22.1%; Pred. No. 0.00011;
Matches 138; Conservative 103; Mismatches 213; Indels 170; Gaps 21;

QY 12 KQNPXSSPG-----QPEAGPEGAQERPSQAAPVAEAGP-----GSSQAPKPEGAQA 60
Db 87 KEENEKSEPKICRNLVTPKADPTG-----SEPAKASEKEAPEDTVDAEGEGRREAAKE 141

QY 61 RTAQSGALRDVSBELSRQLE-----D 81
Db 142 PKKASALEEGSSDAQELISEHMKPEQLSDSIASDPKSFHGLDFGFRSRISEHLLDVD 201

QY 82 ILSTY-----CVDNQGGPGDGAQGPAPPE--DAEKSRITYVANGEP-----EPTPV 128
Db 202 VLSFVLGACRQAQQPLGTEDKDDSDSQDELQSKGLERLSPLPHEERAQSPRS 261

QY 129 VNGEKEPSKGDPNTEBIRQSDVEGD----- 153
Db 262 LATEEPPQGPQEPQEWKEARELIGESASAASLSLQSLQREQAPSPFAACEKGEHQSOAE 321

QY 154 -----RDHRRPQKKVAKGLGKEITLMLQTLTSTPE-----KLAALCKKYAELEHNR 204
Db 322 ELGFGQEAEDPEEKVAVSPPTFPVSPEVRSSTPEVAPPEQLSEALAKAMEEVAQVLE--- 378

QY 205 NSQOMKLLQKQSQOLVQEKHLRGEHSKAVLA--RSKLESICRLQHRNRLKEGVOR 262
```

```

Db 379 --QDORHLESQKQMOOLREKLQOESEEBITLRLHQKQESLSSLRERLQKAEIEBBA-R 435
Qy 263 ARBEEK-----RKEVTSHFQVTLNDIQLOME-----QHNRNSKLRQENNEL 305
Db 436 MREESQRLSWLRAQVQSSTQADEQIRAEQEAASLQKLREBESQQAERASLEQKQRM 495
Qy 306 AERLKKLIEQVELBEHIDKVPKHDLOO---OLVDAKLQQAQEMLXE-----A 351
Db 496 LEQKBEIEASEQEAALNAKAKALQOLREQLEGERKEAVATLEKESAEERLCSL 555
Qy 352 EERHOREKDFLLKEAVESQRMCELMKO---QETHLQKQALALYTEKEFEFQNTL--SKSS 405
Db 556 EAKHREVVSLLQKIQEAQKEBAQLOKCLQGVHVRHVSQYHVGVEHLSLLREKQ 615
Qy 406 EVFTTFQEMKTKKIKLEKETMYRSRWESSNKALLEMAEBKTVRDKLEGLQVKIQ 465
Db 616 EVEGEHERLDM-----KEHQVMAKAREQYE-----AEERKQASILLGLHATGELE 663
Qy 466 RLEKLC-RALQTERNDLNKRVODL 488
Db 564 RLOPAHERLETVRQEQKLEDL 687

RESULT 5
T50985
related to transcription factor TMP [imported] - Neurospora crassa
N:Alternate names: protein B7F18.40
C:Species: Neurospora crassa
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
A:Accession: T50985
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A:Reference number: 225286
A:Accession: T50985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-892 <CH>
A:Cross-references: EMBL:AL389891; GSPDB:GN00116; NCSP:B7F18.40
A:Experimental source: BAC clone B7F18; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F18.40
A:Map position: 6
A:Introns: 63/1; 551/3

Query Match 10.0%; Score 278; DB 2; Length 892;
Best Local Similarity 23.2%; Pred. No. 0.00013;
Matches 142; Conservative 106; Mismatches 247; Indels 118; Gaps 27;

Qy 2 KNDKQKGAQKSNPKSPGQPEAGPEGACERPS---QAAPAVEAGPGSSQAPRKDEGA 58
Db 91 KNLD-RTSSSTQASPRQSMAPASRASTSTSRPNFADKASFNNAVSPRASADTPKRSQDT 149
Qy 59 --QARTAQSGALRDVSBELSRQLEDILSTYCVNNQGGPGEDGAQGPAPDEAKSRTY 116
Db 150 TQEPVPAPLDSADDIKEVAQSVSE-----XADSRPSTD-QPSPENTEDEAPT--- 197
Qy 117 VARGNEPEPTVNGEKEPKSGDNTHEIRQSDVGRDRHRRPOEKKKAGLKEITLTM 176
Db 198 --TESKLEAPAEILAKTEPKSVETVKARPQDQEMKSDTPQCNQDEIYAYVERIDALE 255
Qy 177 QTL-----NTLSTP-----BEKLAALCKKVAELLEERN--SOKO-----MKL 212
Db 256 AKLOYLAREASAAARKALAPAGASAKKLABKDDQQAQLMEGKNLASNEQKLTILKN 315
Qy 213 LQKQSQOLVQEKHLRGEHGAVALRSKLESCLR---ELQHNHNSLKEGVQVAREBEK 269
Db 316 LRKQQAEDKQNGNLKAAKAKADREIENLRKRARHADELEKSNQVNLQKRLDQSQRELYNL 375
Qy 270 RKEVTSHFQVTLNDIQLOMEQHNER-----NSKLRQENMELABELKKLIEQVELREE 321
Db 376 RPEVSK-DTIIARESQIQAKATQADVMAKANDKAREQQQ-----RRIAL-----EE 424
Qy 322 HIDKVPKHDLOQLOVDKALQQAQOEMLKBAERHQREKDF-----LLKEAVESQR- 371

```

```

Db 425 SVEALKIEKNL---MADRAQAQADELRKEAKERAKALELELKABVHNVQESKLEAMRT 481
Qy 372 -----MCELMKQOETHLQKQALALYTEKEFEFQNTLSKSSEVFTTFKQEMK 418
Db 482 RAEAEASSGYTQDSQAKLLAQVET-LQSQYSIASENWQGIETTL--RSRIVNLSKERDE-- 536
Qy 419 TKKIKKLEKETMYRSRWESSNKALLEMAEBKTVRDKLEGLQVKIQRLKELCR-----AL 474
Db 537 -----ALQESDMRRKAREABE---LEBAKTKLPNOEDVESYRSQDLSLKRABEAAL 588
Qy 475 QTERNDLNKRVODLSAGGOSLTDSPERRPEGGAQAPSSPRVTEAPCYPGAP--STEA 532
Db 589 AERADPEKQKQAWEA--EKELIKERERDLQSQG-----NRPR-SWLEGLPGGPFLLKNEG 641
Qy 533 SGQTGPQEPPTSAR 545
Db 642 SGPGSPQLSTAQR 654

RESULT 6
A38194
desmoplakin I - human
N:Contains: desmoplakin II
C:Species: Homo sapiens (man)
C>Date: 04-Mar-1993 #sequence_revision 03-May-1996 #text_change 21-Jul-2000
A:Accession: A38194; A35536; B35536
R:Virata, M.L.; Wagner, R.M.; Parry, D.A.; Green, K.J.
Proc. Natl. Acad. Sci. U.S.A. 89, 544-548, 1992
A:Title: Molecular structure of the human desmoplakin I and II amino terminus.
A:Reference number: A38194; MUID:92115697; PMID:1731325
A:Accession: A38194
A:Molecule type: mRNA
A:Residues: 1-974 <VIR>
A:Note: sequence extracted from NCBI backbone (NCBI:75984, NCBI:75986)
R:Green, K.J.; Parry, D.A.D.; Steinert, P.M.; Virata, M.L.A.; Wagner, R.M.; Angst, B.D.
J. Biol. Chem. 265, 2603-2612, 1990
A:Title: Structure of the human desmoplakins. Implications for function in the desmosome
A:Reference number: A35536; MUID:90153880; PMID:1689290
A:Accession: A35536
A:Molecule type: mRNA
A:Residues: 'R', 927-2677 <GRE>
A:Cross-references: GB:J05211; NID:gl81607; PIDN:AAA35766.1; PID:gl81608
A:Accession: B35536
A:Molecule type: mRNA
A:Residues: 'R', 927-1000, 1600-2677 <GR2>
A:Cross-references: GB:J05211
C:Comment: Desmoplakins I and II, products of single gene, are major proteins of the in
C:Genetics:
A:Gene: GDB:DSP
A:Cross-references: GDB:126564; OMIM:125647
A:Map position: 6pter-6p21
C:Keywords: alternative splicing; cytoskeleton
F:1-2677/Product: desmoplakin I #status predicted <MAT1>
F:1-1000, 1600-2677/Product: desmoplakin II #status predicted <MAT2>

Query Match 9.4%; Score 263; DB 2; Length 2677;
Best Local Similarity 23.4%; Pred. No. 0.0016;
Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 25;

Qy 42 EAEFGSSQAPRKPF-----EGAAQAPTAQS-----GALRDVSELSRQLEDIL 83
Db 1238 QKATGSEVSQKQLEVLQVQVQRTSESVRYKQSLDAAKTIQDKNKEIER-LKQLI 1296
Qy 84 STYCVNNQGGPGEDGAQGPAPDEAKSRV-YVARGNEPEPTPVN-----GEKEPSKG 138
Db 1297 DKETNDR-----KCLEDEARLQRYQVLDQKANSATETINKLRVQEBELTEL 1344
Qy 139 DPNTEBIROSDEVDGRHRRPOEKKKAGLKEITLMTQNTLSTPEEKLAALCKKYAB 198
Db 1345 RIDYERVSGERTVKQDITRFQNSLKELOLQKQ--KVBEELNLRKXTASEDCSKKLEB 1402
Qy 199 LLEEHNNSQKMKL-----LQKQSQ--LVQEKD-----HLR-GHSHKAVLAR 238

```

```

Db 1403 ELEGHRSLSKEQAIKTNLTQLEQASIVKKSSEDLRQQRDVLGDHLEKQRTQBELR 1462
Qy 239 --SKLESCLRELORHNSLKE--EGVQAPRBEKEKKEVTSHFQVTLNDIQLOQEHQ 291
Db 1463 LSSEVEALRQLQEQSVQKLAHNEHFQKAEKSR-----SLNESKIBERL 1512
Qy 292 NERNSKLQENNELAERLKKL-IEQYELR-----BEHIDK-----VFKHKOQQ 334
Db 1513 QSLTENLTKEHLMLEELNLRLEVDLLRGRSEADSKNATILERSQLQISNNRTLEL 1572
Qy 335 QLVDAKLOAQOEMLEKAEERHQREKDFLLKRAVSORMC-ELMKQOETHLKKQALITYEK 393
Db 1573 QGLINDLQRENLQOIEKTKQKALEASNRIOESKQCTQVQVRESLLVKIKVLEQDK 1632
Qy 394 -----FEFQNTLSKSSSVFTTKQEMKTKKIK-LEKETTMYRSWESSKALLEMA 447
Db 1633 ARLOLEDELNRANKSTLEATFTVKORLECEKQIQNDLNQWTKYYSRKEAIRK--IEST 1690
Qy 448 BEKTVRDK-----ELSGLOVKIQRIEKLOR-----ALQTERNDLNKRVQDLISA 490
Db 1691 REKSPREKNSLRFSETERLQAEIKRIEERCRAKLEDSTRETQSQLETSRYQREIDKL-- 1748
Qy 491 GGQGLSTDSGPERPEGRGAQAPSSPRVTEAPC 523
Db 1749 -----RQRPYG-----SHRETQTEC 1763

RESULT 7
trichohyalin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: S28589
R:Fietz, M.J.; Rogers, G.B.
A:Submitted to the EMBL Data Library, December 1992
A:Description: Examination of the gene encoding rabbit trichohyalin.
A:Reference number: S28589
A:Accession: S28589
A:Molecule type: DNA
A:Residues: 1-1407 <PIE>
A:Cross-references: EMBL:Z19092; NID:g1746; PIDN:CAA79519.1; PID:g1747
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root she
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Introns: 46/3
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 9.4%; Score 261.5; DB 1; Length 1407;
Best Local Similarity 23.5%; Pred. No. 0.00097;
Matches 117; Conservative 110; Mismatches 168; Indels 103; Gaps 22;

Qy 42 EARGPSS-----QAPKPGGAQRTAQSGALRDVSEELSRQLEIDILSTYCVDNQGG 94
Db 514 QEQPQGTWRWQLQBEAQRRHTLYAKPQQQLRE-BEELQ-----BKRRQE 561
Qy 95 PGEDGAQGPAPDEPAKSRVTYVARNGEPTPVVNGEKPSPGDPNTEERQSDVEGDR 154
Db 562 RREVTREELQREBEKQR-----RQREKQVRELEELQBEQLADR 604
Qy 155 DRRPQEKKAQGLGKEITLMTLNTLSTPEKLAALCKKYAEALLEE-HRNSQKMKLL 213
Db 605 KLREBEQLQOER---EEERLRQERERKLREBEQL--LRQEQELRQERERKLREBEQL 659
Qy 214 QKKQSQLVQEKDH-LRGEHSKAVLARSKLCSLREBLQHRNLSLKEGVQVAREES---EK 269
Db 660 RREBEQLQREERXUREEQ-----LIQREBEELRQERARKLRREBEQLRQER 716
Qy 270 RKEVTSHFQVTLNDIQLOQEHNRNSKLQENNEL-----AERLKKLIEQYELREHIDK 325
Db 717 EKLREBEQLRREBEQLLQ---ERDRKLREBEQLQSEERERLRQREBEQLRREER-DR 772

```

```

Qy 326 VFHKD-----LQOQLVDKLOAQOEMLEKAE-----ERHQEKDFLLKEA 366
Db 773 KFRBEQLQOERERERLRQERERKLREBEQLQOERERERLRQERERKLREBEQLQOER 832
Qy 367 VESQRM-----ELMKQOETHLKKQALITYEKFEFQNTLSKSSSVFTTKQ 414
Db 833 -EEERLRQERERKLREBEQLQOERERERERERERERERERERERERERERERER 885
Qy 415 MEKWTKKIKL-----EKETTMYRSWESSKALLEMAEKTIVKDEKLEGVQVQRIEKL 470
Db 886 RDRKLREBEQLQOERERERERERERERERERERERERERERERERERERERERER 942
Qy 471 CRALQTS-RNDLNKRVQD 487
Db 943 LRREBEQLRERERARKLRE 960

RESULT 8
H69378
conserved hypothetical protein AF1032 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
C:Accession: H69378
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.B.; Ketchum, K.A.; Dods,
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: H69378
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-886 <KLE>
A:Cross-references: GB:AE001032; GB:AE000782; NID:g2689355; PIDN:AAB90211.1; PID:g26491;
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 9.3%; Score 260; DB 2; Length 886;
Best Local Similarity 24.6%; Pred. No. 0.0007;
Matches 118; Conservative 90; Mismatches 118; Indels 154; Gaps 23;

Qy 108 EDAREKSTYVARNGEPTPVVNGEKPSPGDPNTEERQSDVEGDRHRRPQEKKAG 167
Db 155 ENAWKNLGAIV-----MLEREKERLK-----EFLSQEEQI-----KRQKEKKA-- 194
Qy 168 LKHEITLLMTLNTLSTPEKLAALCKKYAEALLEEHRNSQKMKLLQKQSQLVQEKDHL 227
Db 195 ---EIERISEIKSIESLREKLS-----EVRNLESRLKELEHEKLSRL-----ESL 237
Qy 228 RGEHSKAVLARSKLCSLREBLQHRNLSLKEGVQVAREEKEKEV----- 273
Db 238 RKQESSVLOEVGRGLEEKRLRE---KQKEV-VERIEDLEKKEKVKELPKABRYSILE 293
Qy 274 --TSHFQVTLND-----IQLOQEHNRNSKLQENNELAERLKKLIEQY 317
Db 294 KLASEINQALRDVEKRGDITREAAQTQALKAEDNSKL-EETIKRIELELERERFE 352
Qy 318 -----LREHII--DKVFKHKLQOQLVDKLOAQOEMLEKAE 353
Db 353 KSHRLLETLPKMDRMQGIKAKLEENLTPDKVEKYVDL-----LSKAKEEKEITE 404
Qy 354 RHQR---EKDFL-----LKEAVE---SORMELAKQO---ETHLKKQALITYEKFEFQ 398
Db 405 KLKGLIAKKSLSLTRGAQLKKAVEELKSAERTCPVCGRELDDEHRRKNIMABYTREMRIA 464
Qy 399 NTLKSSESSEVFTTKQEMKTKKIKLEKETTMYRSW-----ESSNKALL 444
Db 465 BELAQADEIKKLERKEVE---KALEKETVTKRWQVDELKALENELSSHDAEKLSA 521
Qy 445 EMAREKTVRDKLEGL--QVKI-----ORLEKLCRALQTERNDLNKRVQD 487

```

Db 522 ESSEYRKVKSR-LDGLRGQQKILLSSASRIKELKSSLIREFEALNKVNSRGELHKKIRE 590

RESULT 9  
T34021  
protein kinase SK2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
R:Accession: T34021  
R:Fukami, Y.; Yamamoto, H.; Ichihara, T.; Mori, K.; Gomi, T.; Sato, K.  
submitted to the EMBL Data Library, April 1997  
A:Description: SK2, a putative rat homologue of yeast protein kinase NRK1.  
A:Reference number: Z21463  
A:Accession: T34021  
A:Status: preliminary; translated from GB/EMBL/DBDJB  
A:Molecule type: mRNA  
A:Residues: 1-1206 <FUK>  
A:Cross-references: EMBL:AB003357; NID:di106513; PID:di020890; PIDN:BAA20077.1  
A:Experimental source: strain SD  
C:Genetics:  
A:Gene: SK2

Query Match 9.3%; Score 258; DB 2; Length 1206;  
Best local Similarity 22.3%; Pred. No. 0.0012;  
Matches 153; Conservative 114; Mismatches 218; Indels 200; Gaps 34;

Qy	2	KNQDKNGAAKQS-----NPKSSPGQAPGPGACERP-----SQAAPAVEA	43
Db	562	QSQDAGGAGEAEPEPAQTLEKATEG-PEA--HGAREEPGSRVEDKQLEQQSAYCEG	618
Qy	44	EGPSSQAPRPEGAQART--AQSGALRDVSEELS-ROLEDILSTYCVDNNGQPG	96
Db	619	EG-----QVTSTSESTRATTEPETDEVDQVSESNIBELERLGVTGAEQALGSKGEAAT	674
Qy	97	-----EDGAGCEP--AEPEDAEXKRT-----VYARNGE-----PEPTPV	128
Db	675	ELDLERENAEQLPVKAEPQAPAAQSAQSEPPVLIPSTNIHSENTENKNGALPKPPTI	734
Qy	129	VNGEKPEKSG--DPNT-----BEIROSDEVGDRDHRPOE--KKK	164
Db	735	LPPFPENGKNDTSGTGVSESSDLNLSSFLSKTKDQSGSVSLQETRRQKTKLKT	794
Qy	165	AKGL--GKEITLIMQTLNTLSTPEKLAALCKVAELLEEHNRSQKQWLLQKQ	217
Db	795	RKFTVDGVGEVSV--TTSKIVTSD-----SKTEELFLRRQBLRLQLLQEEQAKAQ	845
Qy	218	---SOLQEKDHLRGEHSKAVLASK-----LESICRELQRHNSLKEEGVORAREE--	267
Db	846	QLNGKLOQOREQI FRPEQEHLSKRRQYDQEIENLEKQKQTI EELQGEHTNRLDEAKR	905
Qy	268	---EKRREVTSHFQVTLNDIOLQMEQHNRNSKLRQENMELAKRLKLIHQ	315
Db	906	IKGEOEKEL-SKFPQMLRN-----RKKEQEFVQKQOQLDGLAKTKIIQQQKAEALNIE	958
Qy	316	-----YELREEHIDKVFYHKDLOQLVDVAKLQQAQEMLKEAR--ER	354
Db	959	RECINNKOQLLRAREAAIWELEERHLQZ--QHLLKQKQDQYFIQRHOLLRRHKEHQE	1016
Qy	355	HQREKDFLLKEAV--ESQRMCELMKQOETHLKQOLALY-----TEKPEEF	397
Db	1017	MQRYNQRLIELKNRQTCERARLPKIQSSEAKTRWAMFKSLRINSTATPDODREKIQF	1076
Qy	398	QNTLSKSSEVTTTFQMEKMTKKIKLEKETTYMRSWSESNKALLEMAEK-----	450
Db	1077	AAQEEK-----EQKNERMAQH-QKESQVRDLQOCENAVRELHQLEKCHLLVEH	1127
Qy	451	-TVRDKELE--GLQVKIKRLEKLCRALQTERNDINKRVQDLS-----AGQGSITDSGP	501
Db	1128	ETQKLKELDEEHSQELKEWR-EKLPRCKTILEEPARKLQEQEVFKMTGSECLNPSAQ	1186
Qy	502	ERRPEPGCAQAPSSPRVTEAPCYPG	526
Db	1187	SR-----GCLQTSHPSSSTRAAWAG	1206



A;Residues: 1-205 <HAY>  
 R;Komine, Y.; Maita, T.; Matsuda, G.  
 J. Biochem. 110, 60-67, 1991  
 A;Title: The primary structure of skeletal muscle myosin heavy chain: II. Sequence of the  
 A;Reference number: PX0051; MUID:92041768; PMID:1939028  
 A;Accession: PX0051  
 A;Molecule type: protein  
 A;Residues: 206-636 <KOM>  
 R;Maita, T.; Miyaniishi, T.; Matsuzono, K.; Tanioka, Y.; Matsuda, G.  
 J. Biochem. 110, 68-74, 1991  
 A;Title: The primary structure of skeletal muscle myosin heavy chain: III. Sequence of the  
 A;Reference number: PX0052; MUID:92041769; PMID:1939029  
 A;Accession: PX0052  
 A;Molecule type: protein  
 A;Residues: 201-213;632-837 <MAI>  
 R;Maita, T.; Yajima, E.; Nagata, S.; Miyaniishi, T.; Nakayama, S.; Matsuda, G.  
 J. Biochem. 110, 75-87, 1991  
 A;Title: The primary structure of skeletal muscle myosin heavy chain: IV. Sequence of the  
 A;Reference number: JX0178; MUID:92041770; PMID:1939030  
 A;Accession: JX0178  
 A;Molecule type: protein  
 A;Residues: 833-1938 <WA2>  
 R;Maita, T.; Hayashida, M.; Tanioka, Y.; Komine, Y.; Matsuda, G.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 416-420, 1987  
 A;Title: The primary structure of the myosin head.  
 A;Reference number: A26365; MUID:87092420; PMID:3467365  
 A;Accession: A26365  
 A;Molecule type: protein  
 A;Residues: 1-129, 'X', 131-139, 141-550, 'X', 552-754, 'X', 756-784, 'QL', 787-804, 806-810 <MA3>  
 R;Watanabe, B.  
 Biol. Chem. Hoppe-Seyler 370, 55-61, 1989  
 A;Title: Amino-acid sequence of the hinge region in chicken myosin subfragment-2.  
 A;Reference number: S02082; MUID:89228549; PMID:2713098  
 A;Accession: S02082  
 A;Molecule type: protein  
 A;Residues: 1144-1270 <WAT>  
 R;Yajima, E.  
 Nagasaki Igakukai Zasshi 65, 409-430, 1990  
 A;Title: Study on tail region of skeletal muscle myosin; primary structure and protease  
 A;Reference number: PW0009  
 A;Accession: PW0009  
 A;Molecule type: protein  
 A;Residues: 1304-1938 <YAJ>  
 R;Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.  
 submitted to the EMBL Data Library, August 1991  
 A;Description: Analysis of the chicken fast myosin heavy chain family: Localization of  
 A;Reference number: S39081  
 A;Accession: S39081  
 A;Molecule type: mRNA  
 A;Residues: 1081-1203, 'DV', 1206-1342, 'E', 1344-1544, 'S', 1546-1795, 'HV', 1798-1829, 'S', 1831  
 A;Cross-references: EMBL:M74084  
 R;Moore, L.A.; Arrizubieta, M.J.; Tidyman, W.E.; Herman, L.A.; Bandman, E.  
 J. Mol. Biol. 225, 1143-1151, 1992  
 A;Title: Analysis of the chicken fast myosin heavy chain family. Localization of isoform  
 A;Reference number: S24348; MUID:92309413; PMID:1377278  
 A;Accession: S24351  
 A;Molecule type: mRNA  
 A;Residues: 1082-1182, 'T', 1184-1203, 'DV', 1206-1342, 'E', 1344-1544, 'S', 1546-1780, 'M', 1782-  
 A;Cross-references: EMBL:M74084  
 R;Watanabe, B.  
 Biol. Chem. Hoppe-Seyler 370, 1027-1034, 1989  
 A;Title: Complete amino-acid sequence of subfragment-2 in adult chicken skeletal muscle  
 A;Reference number: S05515; MUID:90121764; PMID:2610940  
 A;Accession: S05515  
 A;Molecule type: protein  
 A;Residues: 842-906, 'Q', 908-1270 <WA3>  
 R;Watanabe, B.  
 Biol. Chem. Hoppe-Seyler 370, 549-558, 1989  
 A;Title: Amino-acid sequence of the short subfragment-2 in adult chicken skeletal muscle  
 A;Reference number: S04501; MUID:89374803; PMID:2775482  
 A;Accession: S04501  
 A;Molecule type: protein  
 A;Residues: 852-906, 'Q', 908-1108 <WA2>

R;Matsuda, G.; Maita, T.; Miyaniishi, T.; Hayashida, M.  
 J. Protein Chem. 6, 33-46, 1987  
 A;Title: Structure and function of muscle myosin.  
 A;Reference number: A60877  
 A;Accession: A60877  
 A;Molecule type: protein  
 A;Residues: 1-139;141-205 <MA1>  
 R;Gulick, J.; Kropp, K.; Robbins, J.  
 J. Biol. Chem. 260, 14513-14520, 1985  
 A;Title: The structure of two fast-white myosin heavy chain promoters. A comparative at  
 A;Reference number: A32507; MUID:86033956; PMID:2997212  
 A;Accession: A34124  
 A;Molecule type: DNA  
 A;Residues: 'X', 1-168 <GUL>  
 A;Cross-references: GB:M13512; GB:M12083; GB:M13510; NID:G212363; PIDN:AAA49866.1; PID:  
 R;Kropp, K.; Gulick, J.; Robbins, J.  
 J. Biol. Chem. 261, 6613-6618, 1986  
 A;Title: A canonical sequence organization at the 5'-end of the myosin heavy chain gene  
 A;Reference number: A32587; MUID:86196091; PMID:3009465  
 A;Accession: C25217  
 A;Molecule type: DNA  
 A;Residues: 'M', 1-56, 'T', 58-76, 'I', 78-168 <KRO>  
 A;Cross-references: GB:M13515; GB:M13511; NID:G212373; PIDN:AAA48971.1; PID:9555468  
 C;Comment: This is a fragment of the globular head.  
 C;Superfamily: myosin heavy chain; myosin motor domain homology  
 C;Keywords: acetylated amino end; actin binding; ATP; coiled coil; hydrolase; methylene  
 P;1-1938/Product: myosin heavy chain #status experimental <MAT>  
 P;89-768/Domain: myosin motor domain homology <MYOT>  
 P;179-186/Region: nucleotide-binding motif A (P-loop)  
 P;550-587/Region: actin binding #status predicted  
 P;657-679/Region: actin binding #status predicted  
 P;841-1938/Domain: coiled coil <COI>  
 P;841-1289/Region: S2  
 P;852-1108/Domain: short subfragment 2 <SUB2>  
 P;1290-1938/Region: light meromyosin  
 P;1/Modified site: acetylated amino end (Ala) #status experimental  
 P;35/Modified site: N6-methyllysine (Lys) #status experimental  
 P;130,551/Modified site: N6,N6,trimethyllysine (Lys) #status experimental  
 P;185/Binding site: ATP (Lys) #status predicted  
 P;697,707/Active site: Cys #status predicted  
 P;755/Modified site: 3--methylhistidine (His) #status experimental

Query Match 9.1%; Score 254; DB 1; Length 1938;  
 Best Local Similarity 22.3%; Pred. No. 0.0027;  
 Matches 122; Conservative 83; Mismatches 169; Indels 174; Gaps 19;  
 QY 49 SQAPKPEGAQARTAQSGALRDVSELSRQLEDILSTYVDNNQCGPGEDGAGQGPAPAE 108  
 DB 1090 SOIQSKIEDQLGQ---LQKKIKLQARTEL---EEIEAERTSA 1132  
 QY 109 DAESRTYVARNGEPEPTPVVNGEKEPSKGDPTTEIRQS-DEVG-----DRHRRPQ 160  
 DB 1133 KAEKRAILSR-----ELEISERLEFAGGATAAQIEMKKREA 1171  
 QY 161 EKKKAGLGGKITLMLQTLNTLSTPEKLAALCKV---ALLSEHRNSQOMKLLQKK 216  
 DB 1172 EFQKWRDLDEATL-----QHEATAALRKXHAOSTAELGSCINLQVKKLEKE 1222  
 QY 217 QSQLVQEKDHLGEGHSAVKARSKLESLEL-----QRNRSLEBEGVQAR- 264  
 DB 1223 KSELNMEIDDLASNMESVSKAKANLEKMCRTLEDQLSEIKTKEQNMINDLNTQARL 1282  
 QY 265 -----EHEEKEVTSI-H-FOYTLNDI 284  
 DB 1283 QTETGEYSRQAEKDALISQLSRGQGTQQTQTEELKXRLHEEIKAKNALAHALQSAHDC 1342  
 QY 285 QLQMEQHNER-----NSKLR-----QENMELAEKLLKLEQVELR 319  
 DB 1343 DLLRQYEEQAKGELQALSKANSEVAQETKYETDAIQTEELBEAKKLAQLQDA 1402  
 QY 320 BEHIDKVFHKDLQOOLVDAKLQQQAEMLKEAEERHOREKDFLKEAVESQRMELKKQ 379  
 DB 1403 BEHV-----EAVNAKAS-----LEKTQRIQNEVEDLAVDVERSNAAALDKK 1447

```
QY 380 ETHLQQLALYTEKEFEFQNTL-----SKSSSEVF---TFPKQ---EMKMTKKIKKL 425
D 1418 QNFVKILAEWKQKYEETQTELEASQKESRUSLSTELFNKKNAIYBESLDHLSTLAKRNKL 1507
QY 426 EKETTMYRSWESSNKKALLEMAEKTVRDKEGLQVKIQBLE-----KLCRALQTE 477
D 1508 QOETADLTQIAEGGKAVHELEKVKVHVEQEKSELSQALBEASLEHEGKILR-LQLE 1566
QY 478 RNDLKNRV 485
D 1567 LNQIKSEI 1574

RESULT 14
A56734
ribosome receptor, 180k - dog
C:Species: Canis lupus familiaris (dog)
C>Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 08-Oct-1999
C:Accession: A56734
R:Wanker, E.E.; Sun, Y.; Savitz, A.J.; Meyer, D.I.
J. Cell Biol. 130, 29-39, 1995
A>Title: Functional characterization of the 180-kD ribosome receptor in vivo.
A:Reference number: A56734; MUID:95310363; PMID:7790375
A:Accession: A56734
A>Status: Preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1534 <WAN>
A:Cross-references: GB:G984113; PIDN:CAA60676.1; PID:G984114
C:Keywords: endoplasmic reticulum; membrane protein; protein biosynthesis
F:198-743/Region: 10-residue repeats (N-Q-G-K-A-E-G-A-P)

Query Match 9.1%; Score 253.5; DB 2; Length 1534;
Best Local Similarity 20.7%; Pred. No. 0.0023;
Matches 158; Conservative 115; Mismatches 229; Indels 261; Gaps 31;

QY 3 NQDKKGAAGKQSNPKSSQPEAGPE-----CAORRPSQAA 38
D 697 NQKK-GBCTTNOGKSSGNSQKVDASANQSKRASPLOGKNADWVQSQAPKQEA 755
QY 39 PAVRABGPGSSQAPKP-----EGAAQ 70
D 756 PAKKSGSKGEGEPGPPSDSPLPYKTLVSTVGSVMVFNEGEAQLTIELSEKAGVIQD 815
QY 71 VSEB-----LSROLED---ILSTYCVNNQGGPGEDGAQGEPAEPDEAKSRT-- 115
D 816 THKATQKGDPAVKRQLEKEKLLAT-----EQDAAVAKSL 855
QY 116 -YVANGPEPTPVVNGEKPSPK---GDPNTEBIROSDEYDGRHRRPQEKKAK--- 166
D 856 RVNKLAAEAKAAAGKAVKQKLVAREQBITAVQARIEASYREHVKEVQQLQCKIRTL 915
QY 167 -----GLGKEITILMQTLN-----TLSTPEEKLALCKKYVALL 201
D 916 QEQLNGNTQLARLQQENSILRDALNQTQSVSKONTLEAKLRQELSKVSKELVSE 975
QY 202 EHRNSQMKLLQKKQSVQEKDHLRGEHSKAVLA-RSKLESICREL---QRNRSIKE 257
D 976 AARQEEQKKALETKTAALEKQVLQASHKSEBEALQKRLDEVSRLCSRTSHASLRA 1035
QY 258 EGVQBARBEERKEKREVTSHFQVTLNDIQLQWQHNSKL---RQENMELAEPLK--- 311
D 1036 D-AEKAQEQQQAALHSLKQSSRAVSKSEELSGHLGQKEAPANSQLMIRSI 1094
QY 312 LIEQVELAE-----EHTDKVFKKDLQQLQVDAKLQQAQKMLKEABERHO--- 356
D 1095 LLEAGQARDTQDAQASRAEHQARL---KELBSQV--WCLEKEATELKEAVQCKVKNNDL 1149
QY 357 REKDFLLKEAVES-QRMCELMKQOETHLKQ-----LAL----- 389
D 1150 REKNWKAEMALASARACEEKLRLSTQAKBESKQLSLTEAQTKKALLALLPALSSAPQ 1209
QY 390 -YTE-----KFEFQNTLSKSEVFTTFKQEMEK 417
```

```
D 1210 SYTEMQLQELREKPELLKQRPADTDPSSDLASKLREAEFTQNNLQACDQYRILATEG 1269
QY 418 MTKKIKK-LEKETTMYRSWESSNKKALLEMAEKTVRDKEGLQVKIQ----- 465
D 1270 MLKDLQKSVEEBQV---WKAKVSATSEBLOKSRVTVKHLEDIVKELKGELESSEQVRE 1325
QY 466 -----RLKELCRALQTERNDLKNKVDLSAGGQSLTDSGPPERPEGCAQAPSS--- 515
D 1326 HTSHLEALEKHWAAASAEQSYAKEV---AGLRQLLESQSLDAAKSEAQKSNELA 1381
QY 516 ---PRVTEAPCY-----PCAPSTASGQTGPQEPSTARA 546
D 1382 LVRQLSEMKSHVEDGDVAGSPAAPPAAE-----QDPVELKA 1417

RESULT 15
A40691
trichohyalin - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: A40691; A34209; S32633
R:Pietz, M.J.; McLaughlin, C.J.; Campbell, M.T.; Rogers, G.E.
J. Cell Biol. 121, 855-865, 1993
A>Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-binding
A:Reference number: A40691; MUID:93260018; PMID:7684041
A:Accession: A40691
A:Molecule type: DNA
A:Residues: 1-1549 <EM>
A:Cross-references: EMBL:Z18361; NID:G295940; PIDN:CAA79165.1; PID:G295941
A>Note: sequence extracted from NCBI backbone (NCBI:132511)
R:Pietz, M.J.; Presland, R.B.; Rogers, G.E.
J. Cell Biol. 110, 427-436, 1990
A>Title: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker
A:Reference number: A34209; MUID:90130632; PMID:2298812
A:Accession: A34209
A:Molecule type: mRNA
A:Residues: 1016-1151, 1205-1257, 1281-1398, 'G', 1400-1549 <FI2>
A:Cross-references: GB:X51695; NID:G1827; PIDN:CAA35992.1; PID:G1828
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root &
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Introns: 46/3
A>Note: single copy gene
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <BP2>
F:387-851/Region: 28-residue repeats
F:886-1519/Region: 23-residue repeats

Query Match 9.1%; Score 253.5; DB 1; Length 1549;
Best Local Similarity 22.7%; Pred. No. 0.0023;
Matches 111; Conservative 108; Mismatches 194; Indels 77; Gaps 17;

QY 22 QPAGPEGQERPSQAAPAVEAEGPSSQAPKPEGAQARTAGSGALRDVSELSQLED 81
D 521 QREERKQRERQRYLEKVELQEBEQLOQREKREKQ-----YLERVELQEE 574
QY 82 ILSTYCVNNQGGPGEDGAQGPABEDAEKSRITYVARNGEPEPTPVNNGEKPSPKDPN 141
D 575 QL-----QREKQRERQRYLEKVELQEBEQLOQREKREKQRE 617
QY 142 TE-----EIQSDEVDGRHRRPQEKKAGLGEITLLMQTLNTLSTPEEKLAALCKKY 196
D 618 RQYLEKVELQEBEQVQREKRRQRERQRYLEK-----LQROBERL-----QBE 663
QY 197 AEILLEHRNSQMKLLQKKQSVQVQKDL-RGEHSKAVLARSKLESICRELQENRSL 255
D 664 EQLREERKQRERQRYLEKVELQEBEQLOREBERKQRERQRYLEKBELOQREERL 723
QY 256 KEEGVQBARBEERKEKREVTSHFQVTLNDIQLQWQHNSKL-ERNKSLRQENMELAEPLK 314
D 724 QREKQLQREDEKRRQVRERKYL---BEELQOEEDRLQREKQLLEDREKQYLEKVEL 780
```

```
QY 315 QVELREHHIDKVEKHNDLQOLVDKLOQACQEMKLEAEERHOREKDFLLKEAVESORMCE 374
Db | | : : : : : : : : : : : : : : : : : : : : : : : : |
781 ORB--EEQLORERKRGRERQ-----YREELRLREERLHRKEQQ-LORECECEKRROR 831
QY 375 LMKQOETHLKQOALYTEKFEFF-----QNTLSKSSEVFTTPQEMERMTKIKKLEKE 428
Db | | : : : : : : : : : : : : : : : : : : : : : : : : |
832 LERQLEEEELQRL----DRKRQFRDDDDQHQNEV-RNSRVYSKRENKEKS----RQLD-D 881
QY 429 TTMYRSRWESSNKALLEMAEEKTVRDXELEGLOVK-----IQBLEKLCRALQTERNDLNK 483
Db : : : : : : : : : : : : : : : : : : : : : : : : |
882 SWVRESOFQODLPLQDEQEBKERECEWRSRQKRDSPAPBQLLEREQQKETERDRKPF 941
QY 484 RVQDLSAGGQ 493
Db | : : |
942 RREEQLKGQ 951
```

Search completed: June 8, 2004, 16:35:10  
Job time : 37 secs



GenCore version 5.1.6

Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:27:01 ; Search time 17 Seconds  
(without alignments)

1672.371 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 2785

Sequence: 1 MKNQKKGAKQSNPKSP.....AFSTRASQGTGQEPPTSARA 546

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1315	47.2	259	YL14 HUMAN	P40222 homo sapien
2	267	9.6	1084	MYSS_RABIT	P02562 oryctolagus
3	263	9.4	2871	DESP_HUMAN	P15924 homo sapien
4	261.5	9.4	1407	TRHY_RABIT	P37709 oryctolagus
5	260	9.3	886	RA50_ARCFU	O29230 archaeoglob
6	259	9.3	1410	RRB1_HUMAN	Q9P269 homo sapien
7	257	9.2	1938	MYSS_CHICK	P13538 gallus gall
8	254.5	9.1	880	RA50_PVRAB	Q9U2C8 pyrococcus
9	253.5	9.1	1534	RRB1_CANFA	Q28298 canis faml
10	253.5	9.1	1549	TRHY_SHEEP	P22793 ovis aries
11	253.5	9.1	1690	C190_DROME	Q9V7E5 drosophila
12	249	8.9	1938	MYH4_RABIT	Q28641 oryctolagus
13	248	8.9	1605	RRB1_MOUSE	Q99P15 mus musculus
14	246.5	8.9	4473	PLEB1_CRIGR	Q9J155 cricetus
15	246	8.8	1427	REST_HUMAN	P30622 homo sapien
16	245	8.8	1927	MYH8_HUMAN	P13535 homo sapien
17	244.5	8.8	3259	GOB1_HUMAN	P14789 homo sapien
18	243	8.7	1939	MYH6_HUMAN	P13533 homo sapien
19	243	8.7	1940	CHICK CHICK	P02565 gallus gall
20	242.5	8.7	1898	TRHY_HUMAN	Q07283 homo sapien
21	242.5	8.7	4687	PLEB1_RAT	P30427 rattus norv
22	240.5	8.6	1934	MYH7_MSAU	P13540 mesocricetu
23	240	8.6	2663	CENE_HUMAN	Q02224 homo sapien
24	240	8.6	4684	PLEB1_HUMAN	Q15149 homo sapien
25	239.5	8.6	1939	MYH1_HUMAN	P12882 homo sapien
26	238.5	8.6	993	SCP1_MOUSE	Q02209 mus musculus
27	238	8.5	978	RA50_AQUAE	O67124 aquifex aeo
28	237	8.5	1411	BEAL_HUMAN	P15075 homo sapien
29	236	8.5	1935	MYH7_HUMAN	P12883 homo sapien
30	235	8.4	1938	MYH6_MOUSE	Q02566 mus musculus
31	235	8.4	1938	MYH6_RAT	P02563 rattus norv
32	235	8.4	1938	MYH8_HUMAN	Q9UKX3 homo sapien
33	234	8.4	1935	MYH7_PIG	P79293 sus scrofa

## ALIGNMENTS

## RESULT 1

```

YL14_HUMAN
ID   YL14_HUMAN          STANDARD;          PRT;   259 AA.
AC   P40222; Q8N2Y3;
DT   01-FEB-1995 (Rel. 31, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical protein initially thought to be identical with
DE   interleukin-14 (IL-14) (High molecular weight B-cell growth factor)
DE   (HMW-BCGF).
GN   IL14.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
EX   MEDLINE=93317675; PubMed=8327514;
RA   Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA   Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;
RT   "Identification of a cDNA for a human high-molecular-weight B-cell
RT   growth factor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:6330-6334(1993).
RN   [2]
RP   ERRATUM.
RA   Ambros J.L. Jr., Pippin J., Joseph A., Xu C., Blumenthal D.,
RA   Tamayo A., Claypool K., McCourt D., Srikiatchachorn A., Ford R.J.;
RL   Proc. Natl. Acad. Sci. U.S.A. 93:8154-8154(1996).
RN   [3]
RP   SEQUENCE OF 130-259 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton G., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley A.C., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez R.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002)
CC   -!- CAUTION: The clone described in Ref.1 was initially thought to
CC   code for a high molecular weight interleukin (IL-14). The
CC   translation of this cDNA does not predict the open reading frame

```

```

34 234 8.4 1939 1 MYH6_MSAU P13539 mesocricetu
35 233.5 8.4 1941 1 MYH2_HUMAN Q9UKX2 homo sapien
36 233 8.4 919 1 INCE_HUMAN Q9NGS7 homo sapien
37 231.5 8.3 882 1 RA50_PVRAB P58301 pyrococcus
38 231.5 8.3 1935 1 MYH7_RAT P02564 rattus norv
39 231.5 8.3 2230 1 GOA4_HUMAN Q13439 homo sapien
40 231 8.3 1790 1 USO1_YEAST P23386 saccharomyc
41 230.5 8.3 790 1 BEAL_MOUSE Q8B166 mus musculus
42 229.5 8.2 1938 1 MYH_AEQIR P24733 aequipecten
43 229.5 8.2 2238 1 GOA4_MOUSE Q91VW5 mus musculus
44 229.5 8.2 3911 1 AKA9_HUMAN Q99996 h a-kinase
45 229 8.2 1976 1 MYH8_HUMAN P35580 homo sapien

```

CC for the 60 kDa protein described in Ref.1. A reading frame on the  
 CC plus strand predicts a 7.7 kDa protein. The longest open reading  
 CC frame (shown here) is on the opposite strand and predicts a 36.4  
 CC kDa protein. The relationship of this sequence to IL-14, if any,  
 CC is uncertain.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC -----  
 CC EMBL; L15344; -; NOT ANNOTATED CDS.  
 DR EMBL; BC029686; AAH29686.1; -;  
 DR PIR; A48203; A48203.  
 DR Genew; HGNC:5976; IL14.  
 DR MIM; 147684; -;  
 DR GO; GO:0005576; C:extracellular; NAS.  
 DR GO; GO:0030372; F:high molecular weight B-cell growth factor . . .; NAS.  
 DR GO; GO:0008283; F:cell proliferation; NAS.  
 KW Hypothetical protein.  
 SQ SEQUENCE 259 AA; 30107 MW; 8AB8D1AA6891357C CRC64;

Query Match 47.2%; Score 1315; DB 1; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-44;  
 Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 288 MEQHNRNKLQKQENNELAKRLKLIQYELREHSHIDKVPKHKDQLQQLVDAKLQQAQEM 347  
 Db 1 MEQHNRNKLQKQENNELAKRLKLIQYELREHSHIDKVPKHKDQLQQLVDAKLQQAQEM 60

QY 348 LKEAREHOREKDFLLKEAVESQRCMLKQOETHLXQALALYTKKEFEFQNTLSKSEV 407  
 Db 61 LKEAREHOREKDFLLKEAVESQRCMLKQOETHLXQALALYTKKEFEFQNTLSKSEV 120

QY 408 FTTFQENKEMTKIKKLEKETTWYRSWSSNKALLEWBEKTVRDKEGLQVKIQRL 467  
 Db 121 FTTFQENKEMTKIKKLEKETTWYRSWSSNKALLEWBEKTVRDKEGLQVKIQRL 180

QY 468 EKLCHALQTERNDLNKRVODLSAGQSLTDSGPRRPEGPGACAPSSPRVTEAPCPYGA 527  
 Db 181 EKLCHALQTERNDLNKRVODLSAGQSLTDSGPRRPEGPGACAPSSPRVTEAPCPYGA 240

QY 528 PSTEASGQTGPQEPSARA 546  
 Db 241 PSTEASGQTGPQEPSARA 259

RESULT 2  
 ID MYSS RABIT STANDARD; PRT; 1084 AA.  
 AC P02562;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Myosin heavy chain, skeletal muscle (fragments).  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 CX NCBI\_TaxID=9996;  
 RN [1]  
 RP SEQUENCE OF 1-258.  
 RA Capony J.-P., Elzinga M.;  
 RT "The amino acid sequence of A 34,000 dalton fragment from S-2 of  
 RT myosin.";  
 RL Biophys. J. 33:148A-148A(1981).  
 RN [2]  
 RP SEQUENCE OF 259-428.  
 RX MEDLINE=85131142; PubMed=3972832;  
 RA Lu R.C., Wong A.;  
 RT "The amino acid sequence and stability predictions of the hinge

RT region in myosin subfragment 2.";  
 RL J. Biol. Chem. 260:3456-3461(1985).  
 RN [3]  
 RP SEQUENCE OF 409-1084 FROM N.A.  
 RX MEDLINE=87304245; PubMed=3305014;  
 RA Maeda K., Szakiel G., Wittinghofer A.;  
 RT "Characterization of cDNA coding for the complete light meromyosin  
 RL portion of a rabbit fast skeletal muscle myosin heavy chain.";  
 RL Eur. J. Biochem. 167:97-102(1987).  
 CC -!- FUNCTION: Muscle contraction.  
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
 CC and 2 regulatory light chain subunits (MLC-2).  
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
 CC characteristic for alpha-helical coiled coils.  
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light  
 CC meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be  
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped  
 CC subfragment (S2). Contains 1 myosin-like globular head domain.  
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.

CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

CC -----  
 CC EMBL; X05958; CAA23391.1; -;  
 DR PIR; A02985; A02985.  
 DR PIR; A05280; A05280.  
 DR PIR; S00084; S00084.  
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 KW ATP-binding; Multigene family.  
 FT NON TER 1  
 FT DOMAIN <1>258 ALPHA-HELICAL TAIPIECE (S2).  
 FT NON CONS 258 259  
 FT DOMAIN <259>1084 RODLIKE TAIL (S2 AND LMM DOMAINS).  
 FT DOMAIN 455 1084 COILED COIL (POTENTIAL).  
 FT VARIANT 405 405 L -> V.  
 FT VARIANT 408 408 V -> L.  
 FT VARIANT 421 421 E -> D.  
 FT VARIANT 423 423 S -> G.  
 FT VARIANT 426 426 K -> R.  
 SQ SEQUENCE 1084 AA; 125488 MW; 229CFD69A6E1F7F0 CRC64;

Query Match 9.6%; Score 267; DB 1; Length 1084;  
 Best Local Similarity 22.9%; Pred. No. 0.00095;  
 Matches 137; Conservative 87; Mismatches 209; Indels 166; Gaps 21;

QY 4 QDKNGAAKQSNPKSPGQPEAGPEGAQRPSPQAAPVAEAGPGSSQAPKPGQAQARTA 63  
 Db 172 EDKVNILTK---AKTKLEQVDDLESLEGEKKIRMDLE-----RAKELSG-DLKLA 220

QY 64 QSGALRDVSELSRQLEDILSTYCVNNQGGPGEDGAQGEPAPEPAKSRVTVAR---- 119  
 Db 221 QETSM-DIENDKQQLDEKLKLEFMTNLQ-----SKIDEQALMTNLQRIEEL 267

QY 120 --NGEPEPTPVNGEKEPSKGDPTNTEIROS-DEVG-----DRHRRPQEKKAAGLG 169  
 Db 268 EESEIARASRAKAEKQRPDLSRELSERLEAGGASQAQIEMKKEAEKVERDL 327

QY 170 KEITLDMQTLNTLSTPEEKLAALCKKY----ABLLBEHRNSQMKLLQKKQSLVQEKD 225  
 Db 328 EEATL-----QHEATAAALRKKHADSVBELGQIDNLQKRVKQKLEKSELAKMEID 378

QY 226 HLRGEHSAKVLARSKLESLCKEL-----QSHNRSLSX----- 257  
 Db 379 DLGNMETSXKAGNLEKMKRTLEDQLSEVTKTEEHSQRLINELSAQKARLHTESEFSR 438

QY 258 -----EGVQRAREEERKEVTSHT-FQVTLNDIQLOMEQHNE 293  
DB 439 QLDEKDMWSQLSGGQAFQIQIEGLKRLQLEETKAKSALAHALQSSRRDCDLLRQYEE 498  
QY 294 R-----NSKL-----QENMELAEKLLIROYELREEHIDKVEK 328  
DB 499 EQEAKAELGRAMSKANSEVSQWTKCETDAIOETEELEAKKXLAORLODAEHV----- 553  
QY 329 HKDLOQQLVDAKQQAQEMLEKAEERHQREKDFLLKAEVSESQRCWELMFQOETHLQQLA 388  
DB 604 EWHKYEETQAELEASQKESRSLSSTEVFKVKNAYEESLDHLETKRENKQLQOEISDLTE 663  
QY 435 RWESSNKALLEMAEKTVRDKLEGLQVKIORLE-----KLCRALQTERNDLNKRV 485  
DB 664 QIAESAKHIELEKVKQIDQKSELOALEEAGSLEHEGKIIR-IQELNQVXSEI 721

RESULT 3  
ID DESP HUMAN STANDARD; PRT; 2871 AA.  
AC P15924; Q75993; Q14189; Q9UHN4;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Desmoplakin (DP) (250/210 kDa paraneoplastic pemphigus antigen).  
GN DGP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM DPI).  
RC TISSUE=Fore skin;  
RX MEDLINE=92115697; PubMed=1731325;  
RA Virata M.L.A., Wagner R.M., Parry D.A.D., Green K.J.;  
RT "Molecular structure of the human desmoplakin I and II amino  
RT terminus";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:544-548 (1992).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM DPI).  
RC TISSUE=Fore skin;  
RX MEDLINE=90153880; PubMed=1689290;  
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,  
RA Angst B.D., Nilles L.A.;  
RT "Structure of the human desmoplakins. Implications for function in  
RT the desmosomal plaque";  
RL J. Biol. Chem. 265:2603-2612 (1990).  
RN [4]  
RP ERRATUM.  
RX MEDLINE=90361712; PubMed=2391353;  
RA Green K.J., Parry D.A.D., Steinert P.M., Virata M.L.A., Wagner R.M.,  
RA Angst B.D., Nilles L.A.;  
RL J. Biol. Chem. 265:11406-11407 (1990).  
RN [5]  
RP SEQUENCE OF 2854-2871 FROM N.A.  
RC TISSUE=Skin;  
RX MEDLINE=20062965; PubMed=10594734;  
RA Whitlock N.V., Ashton G.H., Dopping-Hepenstal P.J., Gratian M.J.,  
RA Keane F.M., Eady R.A.J., McGrath J.A.;  
RT "Striate palmoplantar keratoderma resulting from desmoplakin  
RT haploinsufficiency";  
RL J. Invest. Dermatol. 113:940-946 (1999).  
RN [6]  
RP CHARACTERIZATION.  
RX MEDLINE=98012209; PubMed=9348293;

RA Kowalczyk A.P., Bornslaeger E.A., Borgwardt J.E., Palka H.L.,  
RA Dhalwal A.S., Corcoran C.M., Denning M.P., Green K.J.;  
RT "The amino-terminal domain of desmoplakin binds to plakoglobin and  
RT clusters desmosomal cadherin-plakoglobin complexes";  
RL J. Cell Biol. 139:773-784 (1997).  
RN [7]  
RP LIPIDATION.  
RX MEDLINE=98316349; PubMed=9651377;  
RA Marekov L.N., Steinert P.M.;  
RT "Ceramides are bound to structural proteins of the human foreskin  
RT epidermal cornified cell envelope";  
RL J. Biol. Chem. 273:17763-17770 (1998).  
RN [8]  
RP VARIANT ARVD8 ARG-299.  
RX MEDLINE=22285952; PubMed=12373648;  
RA Rampazzo A., Nava A., Malacrida S., Beffagna G., Baucé B., Rossi V.,  
RA Zimbello R., Simonati B., Basso C., Thiene G., Towbin J.A.,  
RA Danieli G.A.;  
RT "Mutation in human desmoplakin domain binding to plakoglobin causes a  
RT dominant form of arrhythmogenic right ventricular cardiomyopathy";  
RL Am. J. Hum. Genet. 71:1200-1206 (2002).  
RN [9]  
RP VARIANTS SPWHS LYS-287 AND CYS-2366.  
RX MEDLINE=21830938; PubMed=11841538;  
RA Whitlock N.V., Wan H., Morley S.M., Garzon M.C., Kristal L., Hyde P.,  
RA McLean W.H.I., Pulkkinen L., Uitto J., Christiano A.M., Eady R.A.J.,  
RA McGrath J.A.;  
RT "Compound heterozygosity for non-sense and mis-sense mutations in  
RT desmoplakin underlies skin fragility/woolly hair syndrome";  
RL J. Invest. Dermatol. 118:232-238 (2002).  
CC -I- FUNCTION: Major high molecular weight protein of desmosomes.  
CC Involved in the organization of the desmosomal cadherin-  
CC plakoglobin complexes into discrete plasma membrane domains and in  
CC the anchoring of intermediate filaments to the desmosomes.  
CC -I- SUBUNIT: Homodimer.  
CC -I- SUBCELLULAR LOCATION: Innermost portion of the desmosomal plaque.  
CC -I- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=DPI; Synonyms=DPI1;  
CC IsoId=P15924-1; Sequence=Displayed;  
CC Name=DPI1; Synonyms=DP2;  
CC IsoId=P15924-2; Sequence=VSP\_005070;  
CC -I- TISSUE SPECIFICITY: Isoform DPI is apparently an obligate  
CC constituent of all desmosomes; Isoform DPI1 resides predominantly  
CC in tissues and cells of stratified origin.  
CC -I- DOMAIN: The N-terminal region is required for localization to the  
CC desmosomal plaque and interacts with the N-terminal region of  
CC plakophilin 1. The C-terminal region interacts with intermediate  
CC filaments.  
CC -I- PTM: Substrate of transglutaminase. Some glutamines and lysines  
CC are cross-linked to other desmoplakin molecules, to other proteins  
CC such as keratin, envoplakin, periplakin and involucrin, and to  
CC lipids like omega-hydroxyceramide.  
CC -I- DISEASE: Defects in DSP are the cause of familial arrhythmogenic  
CC right ventricular dysplasia-8 (ARVD8) [MTM:607450], an autosomal  
CC dominant disorder.  
CC -I- DISEASE: Defects in DSP are a cause of striate palmoplantar  
CC keratoderma II (PPK2, KPP2 or SPPK2), characterized by skin  
CC thickening in the palms (linear pattern) and the soles (island-  
CC like pattern) and flexor aspect of the fingers; and rarely by  
CC abnormalities of the nails, the teeth and the hair.  
CC -I- DISEASE: Defects in DSP are the cause of skin fragility-woolly  
CC hair syndrome (SPWHS) [MIM:607655], an autosomal recessive  
CC genodermatosis characterized by focal and diffuse palmoplantar  
CC keratoderma, hyperkeratotic plaques on the trunk and limbs, and  
CC woolly hair with varying degrees of alopecia.  
CC -I- SIMILARITY: Contains 17 plectin repeats.  
CC -I- SIMILARITY: Contains 2 spectrin repeats.  
CC -I- SIMILARITY: Belongs to the plakoin or cytolinker family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announcement/](http://www.isb-sib.ch/announcement) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M7830; AAB5135.1; -	GO; GO:0005200; P; Structural constituent of cytoskeleton; TAS.	GOBLULAR 1.	1056
EMBL; AL031058; -; NOT ANNOTATED_CDS.	GO; GO:000544; P; Epidermal differentiation; TAS.	CENTRAL FIBROUS ROD DOMAIN.	1057
EMBL; J05211; AAB35766.1; -	InterPro; IPR00101; Plectin repeat.	GLOBULAR 2.	1945
EMBL; AF139065; AAF19785.1; -	InterPro; IPR002017; Spectrin.	SPECTRIN 1.	1946
PIR; A38194; A38194.	Pfam; PF00681; Plectrin.	SPECTRIN 2.	347
PDB; 1LM5; 31-JUL-02.	SMART; SM00250; PLEC; 17.	COILED COIL (POTENTIAL).	447
PDB; 1LM7; 31-JUL-02.	Repeat; Coiled coil; Phosphorylation; Cytoskeleton;	PLECTIN 1.	858
Genew; HGNC:3052; DSP.	Structural protein; Alternative splicing; Disease mutation;	PLECTIN 2.	1018
MIM; 125647; -	3D-structure; Lipoprotein.	PLECTIN 3.	2009
MIM; 607450; -	DOMAIN 1	PLECTIN 4.	2046
MIM; 607655; -	DOMAIN 2	PLECTIN 5.	2084
GO; GO:0005200; P; Structural constituent of cytoskeleton; TAS.	DOMAIN 3	PLECTIN 6.	2122
GO; GO:000544; P; Epidermal differentiation; TAS.	DOMAIN 4	PLECTIN 7.	2163
InterPro; IPR00101; Plectin repeat.	DOMAIN 5	PLECTIN 8.	2198
InterPro; IPR002017; Spectrin.	DOMAIN 6	PLECTIN 9.	2251
Pfam; PF00681; Plectrin.	DOMAIN 7	PLECTIN 10.	2289
SMART; SM00250; PLEC; 17.	DOMAIN 8	PLECTIN 11.	2327
Repeat; Coiled coil; Phosphorylation; Cytoskeleton;	DOMAIN 9	PLECTIN 12.	2365
Structural protein; Alternative splicing; Disease mutation;	DOMAIN 10	PLECTIN 13.	2406
3D-structure; Lipoprotein.	DOMAIN 11	PLECTIN 14.	2456
DOMAIN 1	DOMAIN 12	PLECTIN 15.	2507
DOMAIN 2	DOMAIN 13	PLECTIN 16.	2610
DOMAIN 3	DOMAIN 14	PLECTIN 17.	2648
DOMAIN 4	DOMAIN 15	6 X 4 AA TANDEM REPEATS OF G-S-R-[SR].	2724
DOMAIN 5	DOMAIN 16	INTERACTS WITH FLAKOPHILIN 1 AND JUNCTION	2762
DOMAIN 6	DOMAIN 17	PLAKOGLOBIN.	2824
DOMAIN 7	DOMAIN 18	Omega-hydroxyceramide glutamate ester	1
DOMAIN 8	DOMAIN 19	(Potential).	584
DOMAIN 9	DOMAIN 20	PHOSPHORYLATION (BY PKA) (POTENTIAL).	
DOMAIN 10	DOMAIN 21	Missing (in isoform DP11).	
DOMAIN 11	DOMAIN 22	/FTid=VSP_005070.	
DOMAIN 12	DOMAIN 23	N -> K (in SFMHS).	
DOMAIN 13	DOMAIN 24	/FTid=VAR_015569.	
DOMAIN 14	DOMAIN 25	S -> R (in ARVD8).	
DOMAIN 15	DOMAIN 26	/FTid=VAR_015402.	
DOMAIN 16	DOMAIN 27	R -> C (in SFMHS).	
DOMAIN 17	DOMAIN 28	/FTid=VAR_015570.	
DOMAIN 18	DOMAIN 29	R -> A (in REF. 2).	
DOMAIN 19	DOMAIN 30	D -> R (in REF. 3).	
DOMAIN 20	DOMAIN 31	SV -> RL (in REF. 2).	
DOMAIN 21	DOMAIN 32	5639B5B7CD4690B7 CRC64;	
DOMAIN 22	DOMAIN 33	9.4%; Score 263; DB 1; Length 2871;	
DOMAIN 23	DOMAIN 34	Best Local Similarity 23.4%; Pred. No. 0.0034;	
DOMAIN 24	DOMAIN 35	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 25	DOMAIN 36	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 26	DOMAIN 37	Query Match	
DOMAIN 27	DOMAIN 38	Best Local Similarity	
DOMAIN 28	DOMAIN 39	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 29	DOMAIN 40	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 30	DOMAIN 41	Query Match	
DOMAIN 31	DOMAIN 42	Best Local Similarity	
DOMAIN 32	DOMAIN 43	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 33	DOMAIN 44	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 34	DOMAIN 45	Query Match	
DOMAIN 35	DOMAIN 46	Best Local Similarity	
DOMAIN 36	DOMAIN 47	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 37	DOMAIN 48	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 38	DOMAIN 49	Query Match	
DOMAIN 39	DOMAIN 50	Best Local Similarity	
DOMAIN 40	DOMAIN 51	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 41	DOMAIN 52	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 42	DOMAIN 53	Query Match	
DOMAIN 43	DOMAIN 54	Best Local Similarity	
DOMAIN 44	DOMAIN 55	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 45	DOMAIN 56	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 46	DOMAIN 57	Query Match	
DOMAIN 47	DOMAIN 58	Best Local Similarity	
DOMAIN 48	DOMAIN 59	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 49	DOMAIN 60	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 50	DOMAIN 61	Query Match	
DOMAIN 51	DOMAIN 62	Best Local Similarity	
DOMAIN 52	DOMAIN 63	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 53	DOMAIN 64	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 54	DOMAIN 65	Query Match	
DOMAIN 55	DOMAIN 66	Best Local Similarity	
DOMAIN 56	DOMAIN 67	Matches 134; Conservative 107; Mismatches 194; Indels 138; Gaps 257;	
DOMAIN 57	DOMAIN 68	42 EABGPGSQAAPKPF-----EQAQRTAQS-----GALRDVSELSRQLEDIL 83	
DOMAIN 58	DOMAIN 69	Query Match	
DOMAIN 59	DOMAIN 70	Best Local Similarity	
DOMAIN 60	DOMAIN 71	Matches 134; Conservative 107; Mismatches 194; Indels	

1432	QOKATGEVSORKQOQLEVEFLBQVMTQMRTEESVRYKQSLDDAAKTQDKNKEIER--LKQLI	1490
84	STYCVDDNNQGGFBGDAQGEPAEPEDAESTR--YVARNGEPEPTPVVN-----GEKEPSKG	138
1491	DKETNDR-----KCLEDEANFLQVQYDLQKANGSATSATETINKLVQEOELTFL	1538
139	DPNTEETIQSDVGDHRHRRPOEKKAKGLKEITLLMCTLNTLSTPEEKLAALCKKYAE	198
1539	RIDTIVRSQERTVQDDITRQNSLKELOLQKQ--KVEEELNRLKRTASEDSCKKKLEE	1596
159	LLEHRNSQOKMWL-----LOKQSQO--LVQEKD-----HLR-GEHSKAVLAR	238
1597	ELEGARRSLKEQAIKITMLTQOEASIVKGRSEDDLQQRDVLGDGHLREKORTQEBELR	1656
239	--SKLESICRELQHRNRLSKS-----EGVORAREEEERKEVTSHQVTLNDIQLOMBQH	291
1657	LSSEVEALRRQLQEQBSVQKQHLNHFQKAIEDKSR-----SLNESKIEIERL	1706
252	NERNSKLQENNELAERLKKU--IEOYELR-----BEHIDK-----VPKHDLQOQ	334
1707	QSLTENLTKEHLMLLEEELNRLRLVEDYDLRRGRSEADSDKNATILELRSQLQISNNRTLEL	1766
335	QIVDAKQQAQOEMLKEASERHOREKDFTLKEAVESQRCM--ELMKQOETHLKQOALYATEK	393
1767	QGLINDLQRENLFQOETEKFOKQALEASNRIQESKNQCTQVVOERESLVAIKVLEQDK	1826
394	-----FEFQNTLSKSSEVPTTFQOEMKMTKKIK--LEKETMTYRSWESSNKALLEMA	447
1827	ARLQRLDELNRKASTLEAFETRVQKRLCEKQOQIQLNDLQNMWKTQYSRKEEAIRK--IESE	1884
448	EKTVRDK-----ELEGQVKIORLEKLCR-----ALOTERDNLNKRQVDLSA	490
1895	REKSERENNSLSNEIERLOAEIKRIEERCKRKLEDSTRETQSLFERSRYOREIDKL--	1942
491	GGQSLTDSGPERRPEGFGAOPSSPRVTEAPC	523
1943	-----RORPYG-----SHRETOTEC	1957

```

RESULT 4
TRHY_RABIT
ID TRHY_RABIT STANDARD; PRT; 1407 AA.
AC P37769;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update);
DE Trichohyalin.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Matheria; Lagomorpha; Leporidae; Oryctolagus.
NCBI_TaxID=9986;
RN [1]
RN SEQUENCE FROM N.A.
RA Pietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: Intermediate filament-associated protein that associates
CC in regular arrays with keratin intermediate filaments (KIF) of the
CC inner root sheath cells of the hair follicle and the granular
CC layer of the epidermis. It later becomes cross-linked to KIF by
CC isodipeptide bonds. It may serve as scaffold protein, together
CC with involucrin, in the organization of the cell envelope or even
CC anchor the cell envelope to the KIF network. It may be involved in
CC its own calcium-dependent postsynthetic processing during terminal
CC differentiation.
CC -|- SUBUNIT: Homodimer (Probable).
CC -|- TISSUE SPECIFICITY: Found in the hard keratinizing tissues such as
CC the inner root sheath (IRS) of hair follicles and medulla, and in
CC the filiform papillae of dorsal tongue epithelium (Probable).
CC -|- DEVELOPMENTAL STAGE: Expressed during late differentiation of
CC the epidermis.
CC -|- DOMAIN: Consists of nine domains. Domain 1 contains two EF-hand
CC calcium-binding domains. Domains 2-4, 6, and 8 are almost entirely

```



QY 108 EDAKSRVTVARNGEPPTFVNGEKEPSKGDPTNTEERISQSDVDRDRRPPQKKKAG 167  
 Db 155 ENAMKNGAVIR-----MLREKERLK-----EFLSQBEQI-----KKQKBEKA-- 194  
 QY 168 LGKEITLLMOTINTLSTPEEKLAALCKKYAEELLEHRNSQKMLLKQKQKQKQKQK 227  
 Db 195 ---EIERISEBIKSIESREKLS-----BEVRNLESLKLEBEHKSRL-----ESL 237  
 QY 228 RGEHSKAVLASKSLCSRELOHNRSLKBEQVQBARBEKREKVE----- 273  
 Db 238 RKQESSVLQVGRGLEKLELE---KQKEV-VERIEDLEKKAKEVVKELKPAERYSILE 293  
 QY 274 ---TSHFQVTLND-----IQLOEQHNRNSKLRQENMELAELKLLKLEOYE 317  
 Db 294 KLLSEINQALRDVEKRGDGLTEAAGIQALKKAEDNSKL-EHITTKRIBELERLEKPE 352  
 QY 318 -----LREHI---DKVFKHDLQOOLVDKALQQAQEMLKAAEE 353  
 Db 353 KSHRLLETLPKMDRMQGIKALBEKNLTPDKVKMYDL-----LSKAKBEKEITE 404  
 QY 354 RHQR---EKDFL-----LKEAVE---SQRMCELMKQO---ETHLKOQALALYTEKPEEFQ 398  
 Db 405 KKLKLIKSSIKTEGAQKAKAVEELKSAERTCPVCGRELEDEKRNIMAYTTEMKRIA 464  
 QY 399 NTLKSSSVFTTFOEMERMTKIKLKEKTTMVRSRW-----ESSNKALL 444  
 Db 465 BELAKADEIEKKLERLEKVE---KALEKQETVLKRYQWDELKALENELSHDAEKL 521  
 QY 445 EMABEKTVRDKELEGL---OVKI-----ORLEKLCALQOTRENDLANKVQD 487  
 Db 522 ESEYRKVKER-LOGLRQOQKILLSSASRIKELKSSUREIBALKNVSEGEUHRKIRE 580

RESULT 6  
 RRB1\_HUMAN  
 ID RRB1\_HUMAN STANDARD; PRT: 1410 AA.  
 AC Q9P2E9; Q75300; Q75301; Q96SB2; Q9BWP1; Q9H476;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Ribosome-binding protein 1 (Ribosome receptor protein) (180 kDa  
 DE ribosome receptor homolog) (ES/130 related protein).  
 GN RRB1 OR KIA1398.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX MEDLINE=98290552; PubMed=9628588;  
 RA Langley R., Leung E., Morris C., Berg R., McDonald M., Weaver A.,  
 RA Parry D., Ni J., Su J., Gentz R., Spurr N., Krissansen G.W.;  
 RT "Identification of multiple forms of 180-kDa ribosome receptor in  
 RT human cells."  
 RL DNA Cell Biol. 17:449-460(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RC TISSUE=Brain;  
 RX MEDLINE=20181126; PubMed=10718198;  
 RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes.XVI.  
 RT The complete sequences of 150 new cDNA clones from brain which code  
 RT for large proteins in vitro."  
 RL DNA Res. 7:65-73(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Kibler A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Leheslahti M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie L.J., McLeay K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.O.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,  
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [4]  
 RP SEQUENCE OF 413-1208 FROM N.A. (ISOFORM 3).  
 RC TISSUE=Uterus;  
 RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 756-1410 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueudin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Vallal D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Acts as a ribosome receptor and mediates interaction  
 CC between the ribosome and the endoplasmic reticulum membrane (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic  
 CC reticulum (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Comment=Additional isoforms seem to exist. Additional isoforms  
 CC are probable deriving from alternative splicing in the repeat  
 CC region;  
 CC Name=3;  
 CC IsoId=Q9P2E9-1; Sequence=Displayed;  
 CC Name=1; Synonyms=pi180;  
 CC IsoId=Q9P2E9-2; Sequence=VSP\_003950;  
 CC Name=2; Synonyms=ES130;  
 CC IsoId=Q9P2E9-3; Sequence=VSP\_003949, VSP\_003950;  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



RP SEQUENCE OF 838-1938.  
 RC TISSUE=Pectoralis muscle;  
 RX MEDLINE=92041770; PubMed=1939030;  
 RA Maizumi T., Tajima E., Nagata S., Miyawaki T., Nakayama S., Matsuda G.,  
 RT "The primary structure of skeletal muscle myosin heavy chain: IV.  
 RT Sequence of the rod, and the complete 1,938-residue sequence of the  
 RT heavy chain.";  
 RL J. Biochem. 110:75-87(1991).  
 RN [6]  
 RP PRELIMINARY SEQUENCE OF 1-808.  
 RX MEDLINE=87092420; PubMed=3467365;  
 RA Maizumi T., Hayashida M., Tanioka Y., Komine Y., Matsuda G.,  
 RT "The primary structure of the myosin head.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 84:416-420(1987).  
 RN [7]  
 RP SEQUENCE OF 842-1270.  
 RX MEDLINE=90121764; PubMed=2610940;  
 RA Watanabe B.;  
 RT "Complete amino-acid sequence of subfragment-2 in adult chicken  
 RT skeletal muscle myosin.";  
 RL Biol. Chem. Hoppe-Seyler 370:1027-1034(1989).  
 RN [8]  
 RP SEQUENCE OF 852-1108.  
 RX MEDLINE=89374803; PubMed=2775482;  
 RA Watanabe B.;  
 RT "Amino-acid sequence of the short subfragment-2 in adult chicken  
 RT skeletal muscle myosin.";  
 RL Biol. Chem. Hoppe-Seyler 370:549-558(1989).  
 RN [9]  
 RP SEQUENCE OF 1145-1270.  
 RX MEDLINE=89228549; PubMed=2713098;  
 RA Watanabe B.;  
 RT "Amino-acid sequence of the hinge region in chicken myosin  
 RT subfragment-2.";  
 RL Biol. Chem. Hoppe-Seyler 370:55-61(1989).  
 RN [10]  
 RP SEQUENCE OF 1857-1938 FROM N.A.  
 RX MEDLINE=87217964; PubMed=3034534;  
 RA Moriarty D.M., Barringer K.J., Dodgson J.B., Richter H.B.,  
 RT Young R.B.;  
 RT "Genomic clones encoding chicken myosin heavy-chain genes.";  
 RL DNA 6:91-99(1987).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 1-843.  
 RX MEDLINE=93303624; PubMed=8316857;  
 RA Rayment I., Rypniewski W.R., Schmidt-Base K., Smith R.,  
 RA Tomchick D.R., Benning M.M., Winkelmann D.A., Wesenberg G.,  
 RA Holden H.M.;  
 RT "Three-dimensional structure of myosin subfragment-1: a molecular  
 RT motor.";  
 RL Science 261:50-58(1993).  
 CC -!- FUNCTION: Muscle contraction. Myosin is a protein that binds to  
 CC F-actin and has ATPase activity that is activated by F-actin.  
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
 CC and 2 regulatory light chain subunits (MLC-2).  
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
 CC characteristic for alpha-helical coiled coils.  
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light  
 CC meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be  
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped  
 CC subfragment (S2).  
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.  
 CC -!- SIMILARITY: Contains 1 IQ domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 -----  
 CC EMBL; U82231; AAB47555.1; -;  
 CC EMBL; M16557; AAA48970.1; -;  
 DR PDB; 2MYS; 11-JAN-97.  
 DR PDB; 1ALM; 17-DEC-97.  
 DR PDB; 1M8Q; 10-SEP-02.  
 DR InterPro; IPR000048; IQ region.  
 DR InterPro; IPR001609; Myosin head.  
 DR InterPro; IPR004009; Myosin\_N.  
 DR InterPro; IPR002928; Myosin\_tail.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF00063; myosin head; 1.  
 DR Pfam; PF02736; Myosin\_N; 1.  
 DR Pfam; PF01576; Myosin\_tail; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR PRODOM; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS00096; IQ; 1.  
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
 KW ATP-binding; Methylation; Phosphorylation; Acetylation;  
 KW Calmodulin-binding; Multigene family; 3D-structure.  
 FT INIT MET 0 0 MYOSIN HEAD-LIKE.  
 FT DOMAIN 1 782 IQ.  
 FT DOMAIN 783 812 HINGE.  
 FT DOMAIN 838 840 COILED COIL (POTENTIAL).  
 FT NP\_BIND 841 1938 ATP (POTENTIAL).  
 FT DOMAIN 179 186 ACTIN-BINDING.  
 FT DOMAIN 657 679 ACTIN-BINDING.  
 FT DOMAIN 759 773 ACETYLATION.  
 FT MOD\_RES 1 1 METHYLATION (MONO-).  
 FT MOD\_RES 35 35 METHYLATION (MONO-).  
 FT MOD\_RES 130 130 METHYLATION (TRI-).  
 FT MOD\_RES 551 551 METHYLATION (TRI-).  
 FT MOD\_RES 755 755 METHYLATION (MONO-).  
 FT CONFLICT 907 907 C -> Q (IN REF. 7 AND 8).  
 FT CONFLICT 980 980 L -> F (IN REF. 1).  
 FT CONFLICT 1343 1343 E -> D (IN REF. 5).  
 FT CONFLICT 1545 1545 S -> A (IN REF. 5).  
 FT CONFLICT 1796 1797 HV -> QL (IN REF. 5).  
 FT CONFLICT 1830 1830 S -> A (IN REF. 5).  
 FT CONFLICT 1863 1863 I -> V (IN REF. 10).  
 FT CONFLICT 1929 1931 IHG -> FH (IN REF. 10).  
 FT TURN 7 8  
 FT TURN 11 12  
 FT TURN 14 17  
 FT HELIX 21 24  
 FT TURN 25 28  
 FT HELIX 34 36  
 FT TURN 37 41  
 FT STRAND 47 55  
 FT STRAND 59 62  
 FT STRAND 70 73  
 FT HELIX 74 76  
 FT STRAND 77 78  
 FT TURN 83 84  
 FT STRAND 90 90  
 FT STRAND 91 93  
 FT HELIX 99 108  
 FT TURN 109 113  
 FT STRAND 116 118  
 FT STRAND 123 126  
 FT HELIX 133 135  
 FT TURN 137 138  
 FT TURN 139 142  
 FT TURN 143 145  
 FT TURN 148 149  
 FT HELIX 155 169  
 FT TURN 170 170  
 FT STRAND 173 179  
 FT TURN 181 182  
 FT HELIX 185 199



```

FT TURN 200 200
FT HELIX 218 233
FT STRAND 234 235
FT STRAND 243 244
FT STRAND 247 254
FT STRAND 260 262
FT STRAND 265 268
FT HELIX 272 275
FT TURN 276 276

Query Match 9.2%; Score 257; DB 1; Length 1938;
Best Local Similarity 22.3%; Pred. No. 0.004;
Matches 122; Conservative 168; Indels 174; Gaps 19;

QY 49 SQAPRPEGAQTAQSGALRDVSELSRLESLSTYCVNNQGGEGGAGGPAEPE 108
DQ 1090 SQQSKIEDQALGMQ---LQKIKELQARIEL-----EEIEARTSRA 1132
QY 109 DAKSRITYVARNCEPEPTVWNGEKPSKDPNTEIRQS-DEVG-----DRDHRPQ 160
DQ 1133 KAEKGRADLR-----SLEISERLEBAGGATAAQIEMNKKREA 1171
QY 161 EKKKAGLGKEITLLMOTLNTLSTPEKLAALCKY-----AELLBEHNSQKMLQKK 216
DQ 1172 EFQMRDLSEATL-----CHEATAAALAKKADSTAEELGEQIDNLQVKKLEKE 1222
QY 217 QSLVQEKDLHGEHSHKAVLARSKLESICREL-----QRHNSLKEGVORAR- 264
DQ 1223 KSELKWEIDDLASNMESVSKANLENKXCTLDQSLSEIKTKEQVQRMINDLQARL 1282
QY 265 -----EEBPKKEVTSH-FQVTLNDI 284
DQ 1283 QTETGEYSRQAEKDALISQSRGKQFTQIBELKRLHEEIKAKNALAHALQSAHDC 1342
QY 285 QLOMEQHNER-----NSKLR-----QENMELAEKLLIEQVELR 319
DQ 1343 ELUREQEEERQAEKGELOALSKANSEVAQWRKYETDAIQRTTELEHAKKQAQLQDA 1402
QY 320 EEHIDKVFKEKLOQQVDAKLOQAQEMKEABERHQEKDFLLKEAVESQRMELMKQ 379
DQ 1403 EEHV-----EAVNAKAS-----LEKTRQLQNEVEDLWVDVRSNAACAALDKK 1447
QY 380 ETHLKQALAYTEKPEFQNTL-----SKSEVP-----TTFKQ-----EMEKMTKIKKL 425
DQ 1448 QKNFDKILAEWKQYESTQTELEASQKESLSLTETLPMKNAYESLHLETKRENKL 1507
QY 426 EKETMYRSWESSNKALLEMAEKTVDKLEGLQVKIQLE-----KLCRALQTE 477
DQ 1508 QQETADLTQIABGGKAVHELEKVKHVQEKSLQASLEEAASLEHBEGKILR-IQL 1566
QY 478 RNDLNKVV 485
DQ 1567 LNOIKSEI 1574

RESULT 8
RASO PYRAB STANDARD; PRT; 880 AA.
AC Q9UUC8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PYRAB12200 OR PAB0812.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS5 / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,

```

```

RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC !- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (BY
CC similarity).
CC !- SUBUNIT: Forms a complex with mre11 (BY similarity).
CC !- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ248286; CAB50131.1; -.
CC F1R; F75103; F75103.
CC HAMAP; MF_00449; -. 1.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003439; ABC transporter.
CC InterPro; IPR007517; Rad50 zn_book.
CC InterPro; IPR003395; SMC_N zn_book.
CC Pfam; PF04423; Rad50 zn_book; 1.
CC Pfam; PF02463; SMC_N; 1.
CC ProDom; PD000006; ABC transporter; 1.
CC SMART; SM00382; AAA; 1.
CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 144 745 COILED COIL (POTENTIAL).
SQ SEQUENCE 880 AA; 103970 MW; F0B177EC7E026479 CRC64;

Query Match 9.1%; Score 254.5; DB 1; Length 880;
Best Local Similarity 23.0%; Pred. No. 0.023;
Matches 132; Conservative 104; Mismatches 206; Indels 133; Gaps 25;

QY 3 NOD---KNGAAKQNPSSP--GOPEAGPEGAQRPSQAAPAVEAGPGSQAPRKEG 57
DQ 202 NDEDTKLKSEINSEISSKLPPIRGLEKVENNVKLESIKGISLKIQLKVKGRK-KG 260
QY 58 AQARTAQSGALRDVSELSR--QLEDILSTYCVNNQGGEGGAGGPAEPEDAEKSR 115
DQ 261 LEEKIVQ---IERSIEKKAKISELEIVKDI-----PKLQEKKE 298
QY 116 YVARG-----EPEPTVWNGEKPSKDPNTEIRQS-DEVGDR- 154
DQ 299 YRKLKGFDRDEYSEKLRLEKELSKWSELSKATSEVKEGKKKRAEIRKLEKRL 358
QY 155 DHRP--QEKKAAGLGKEITLLMOTLNTLSTPE--EKLAALCKKYAEELLEHRNSQKM 210
DQ 359 BELKPYVELEDAKQVQKQIERLKLKLSGVEIKLESLEKERTET-----EEAI 411
QY 211 XLQKQSQQLVQEKHLRGHSHKAVLARSKLSLSC-----REL-QRHRSLK----- 256
DQ 412 KEITTRIGMEQEKN---ERMKAIEELFKAKGKFCVCGRELTTEHKKELMERYTLEIKK 467
QY 257 -EGVQRAAREBEERKEVTSHEFQVTLNDIQLQ---MEQHNRNSKLRSQNMELAE--- 307
DQ 468 IEELKRTTEERKLVNLRKLEIKLREFSVNRDIAEQIKELSKLKGFLNLELEQERE 527
QY 308 -----RLKGLIEQVELBEHIDKVFVK-KDIQQQLVDAKLQ 343
DQ 528 FEGLNPEFNKGLGLGLRLDLKRLKALEGRKLEKVKRAKELEENLRQLRELGPES 597
QY 344 AQEM---LKEAEERHQEKDFLLKEAVESQRMELMKQOETHLKKQALYLYTEKFEONT 400
DQ 588 VVELNLRIOLEEFHDKYVVEAKKSES-ELRELKLNKLEKSEKTELDQAFEMADVENEIEK 646

```

QY	401	LSKSSEVPTTF-KQEMKMTKIKKLEKTTWYSSWESSNKA-----LLWREKTV 452
Db	647	EAKLDESFPSEYERKRLVKLEKREVSLLTLEELKKSVEQIKATLRKLEKEE 706
QY	453	RDKEGLQVKIQLEKLCRALQTERNDLNKRVOD 487
Db	707	REK-----AKLEIKKLEKALSKEV-----DLRKKIND 733
RESULT 9		
RRB1	CANFA	
ID	RRB1	CANFA
AC	Q28298;	STANDARD; PRT; 1534 AA.
DT	28-FEB-2003	(rel. 41, Created)
DT	28-FEB-2003	(rel. 41, Last sequence update)
DT	10-OCT-2003	(rel. 42, Last annotation update)
DE	Ribosome-binding protein 1 (180 kDa ribosome receptor) (RRP).	
GN	RRBP1	OR P180.
OS	Canis familiaris	(Dog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.	
OX	NCBI_TaxID=9615;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.	
RC	TISSUE=Kidney;	
RX	MEDLINE=95310363;	PubMed=7790375;
RA	Wanker E.E., Sun Y., Savitz A.J., Meyer D.I.;	
RT	"Functional characterization of the 180 kDa ribosome receptor in vivo."	
RL	J. Cell Biol.	130:29-39(1995).
CC	- - FUNCTION: Acts as a ribosome receptor and mediates interaction between the ribosome and the endoplasmic reticulum membrane.	
CC	- - SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic reticulum.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.ebi.ac.uk/announcements/">http://www.ebi.ac.uk/announcements/</a> or send an email to <a href="mailto:license@ebi.ac.uk">license@ebi.ac.uk</a> ).	
EMBL	X87224;	CAN60676.1; -.
DR	FIR;	A56734; A56734.
DR	GO;	GO:0030176; C:integral to endoplasmic reticulum membrane; IDA.
DR	GO;	GO:0030376; F:ribosome receptor activity; IDA.
DR	GO;	GO:0007165; P:signal transduction; IDA.
DR	InterPro;	IPR007794; Rib recp KP reg.
DR	Pfam;	PF05104; Rib recp KP reg. I.
KW	Transport; Protein transport; Translocation; Endoplasmic reticulum; Transmembrane; Repeat; Alternative splicing.	
FT	DOMAIN	1 LUMENAL (POTENTIAL).
FT	TRANSMEM	8 28 POTENTIAL.
FT	DOMAIN	29 1534 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	197 736 54 X 10 AA APPROXIMATE TANDEM REPEATS OF N-Q-G-K-K-A-E-G-A-P.
FT	SEQUENCE	1534 AA; 164586 MW; B343BCF12656F3C5 CRC64;
Query Match		9.1%; Score 253.5; DB 1; Length 1534;
Best Local Similarity		20.7%; Pred. No. 0.0043;
Matches 158;		Conservative 115; Mismatches 229; Indels 261; Gaps 31;
QY	3	NODKNGAAGKQSNPKSSPGQEPAGE-----GAQERPSQAA 38
Db	697	NOGKK-GEFTNQKKSSESPNQKGVASANQSKRAEAPIQGNADMVQSEAPKQEA 755
QY	39	PAVEAEGPGSSQAPKRP-----EQAQAR-----TAQSGALRD 70
Db	756	PAKKISGKKKGKGGPPGPPDSPLYPKTLTVGSMVFNQGEAQRILSEKAGVTQD 815
QY	71	VSEE-----LSRQLED-----ILSTVCVDNNQGGPGDGAQGEPAEPDAEKSR-- 115

Db	816	TWHKATQKGDPAVAILKRLQLEKEKLLAT-----BQDAVAASKL 855
QY	116	-VYANGSEPTFPVNGSEKPSK---GDPTEEIRQSDEVDGRHRRPOEKKKAK--- 166
Db	856	REVNKELAAKAKAAGBAKVKQVLVAREQETAVQARIEASYREHVKEVQLOGKIRTL 915
QY	167	-----GLGKEITLLMOTLN-----TLSTPEKLAALCKKYAELE 201
Db	916	OEQLNGPNTQLARLQOENSILRDALNQATSVQESQNTLAKLROELSKVSKELVSE 975
QY	202	EHRNSQKMLKQKQSOLVQEKDHLRGHESKAVLA-RSKLSLQREL---ORHNSLKE 257
Db	976	AARQEEQQRKALETKTAALKQVLOQASHKSEELQRLDEVSELCSQTSHASLRA 1035
QY	258	EGVQARAREEERKKEVTSHFQVTLNDIQLQMEQHNRNSKL---ROENMELAEK-- 311
Db	1036	D-AEKAQEQQQQMAELHSLKSSAEVSKSELSGLHGQLKEARASQLMERIRSIEA 1094
QY	312	LIEQYELRE-----EHIDKVFHKDLOQLVDKLAQQAQEMLKAEERHQ----- 356
Db	1095	LLEAGQARDTQDAQASRAEHQARL---KELESQV---WCLKEKATELKEAVEQOKVKNNDL 1149
QY	357	REKDFLLKEAVES-QRMCELMKQOETHLQQ-----LAL----- 389
Db	1150	REKNWAMEALASAEACCEKLSLQAKESEKQSLTEAQTKKALLPALSSSAPQ 1209
QY	390	-YTE-----KPEFQNTLSKSSSEVFTFKQEMEK 417
Db	1210	SYTENQELREKGPPELLKQRPADTPFSSDLASKLREABETQNNLQACQYRTILATEG 1269
QY	418	MTKKIKK-LEKETTMYRSRWSSNKALLMAEKTVRDKLEGLQVKIQ----- 465
Db	1270	MLADLOKSVEEBQV---WKAKVSATEBELQKSRVTVKHELDIVKLGELSSSQVRE 1325
QY	466	-----RIEKCRALOTRNDLNKRVQDLSAGGQSLTDSGPERPEGCAQAPS--- 515
Db	1326	HTSHLEAELEKHMMAASAECCQSYAKEV---AGLRQLLESQSLDAAKSEAQKSNELA 1381
QY	516	---PRVTEAPCY-----PGAPSTEASGQTGPQBPSTARA 546
Db	1382	LVRQQLSEKMSHVEDGDVAGSPAAPAE-----QDPVELKA 1417
RESULT 10		
TRHY	SHEEP	
ID	TRHY	SHEEP
AC	P22793;	STANDARD; PRT; 1549 AA.
DT	01-AUG-1991	(Rel. 19, Created)
DT	01-OCT-1994	(Rel. 30, Last sequence update)
DT	10-OCT-2003	(Rel. 42, Last annotation update)
DE	Trichohyalin.	
GN	THH.	
OS	Ovis aries	(Sheep).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC	Bovidae; Caprinae; Ovis.	
OX	NCBI_TaxID=9940;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=93260018;	PubMed=7684041;
RA	Pietz M.J., McLaughlan C.J., Campbell M.T., Rogers G.E.;	
RT	"Analysis of the sheep trichohyalin gene: potential structural and calcium-binding roles of trichohyalin in the hair follicle."	
RL	J. Cell Biol.	121:855-865(1993).
RN	[2]	
RP	SEQUENCE OF 1016-1549 FROM N.A.	
RC	STRAIN=Merino-Dorset horn X Border Leicester; TISSUE=Wool follicles;	
RX	MEDLINE=90130632;	PubMed=2298812;
RA	Pietz M.J., Presland R.B., Rogers G.E.;	
RT	"The cDNA-decoded amino acid sequence for trichohyalin, a differentiation marker in the hair follicle, contains a 23 amino acid repeat."	



RESULT 11  
 C190\_DROME STANDARD; Q8INY8; Q8MSD0; PRT; 1690 AA.  
 AC Q9VUE5; Q44929; Q8INY8; Q8MSD0;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Restin homolog (Cytoplasmic linker protein 190) (Microtubule binding  
 protein 190) (d-CLIP-190).  
 GN CLIP-190 OR CQ5020.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM A), FUNCTION, SUBCELLULAR LOCATION, AND  
 RP TISSUE SPECIFICITY.  
 RC STRAIN=Oregon-R; TISSUE=Embryo, and Ovary;  
 RX MEDLINE=98139549; PubMed=9472041;  
 RA Lantz V.A., Miller K.G.;  
 RT "A class VI unconventional myosin is associated with a homologue of a  
 RT microtubule-binding protein, cytoplasmic linker protein-170, in  
 RT neurons and at the posterior pole of Drosophila embryos.";  
 RL J. Cell Biol. 140:897-910(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein J., Brodtier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke K., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,  
 RA Fester C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Martell B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN [3]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Murgall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC STRAIN=Berkely; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A Drosophila full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 RN [5]  
 RP FUNCTION, SUBUNIT, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.  
 RC MEDLINE=20530668; PubMed=11076973;  
 RA Sisson J.C., Field C., Ventura R., Royou A., Sullivan W.;  
 RT "Java lamp, a novel peripheral Golgi protein, is required for  
 RT Drosophila melanogaster cellularization.";  
 RL J. Cell Biol. 151:905-918(2000).  
 CC -1- FUNCTION: Together CLIP-190 and jar may coordinate the interaction  
 CC between the actin and microtubule cytoskeleton. May link endocytic  
 CC vesicles to microtubules. May play a role in formation of furrows  
 CC during cellularization.  
 CC -1- SUBUNIT: Interacts with Lva.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; microtubule-associated. Lva-  
 CC CLIP-190 complexes are found at the Golgi.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=3;  
 CC Name=A;  
 CC IsoId=Q9VUE5-1; Sequence=Displayed;  
 CC Name=B;  
 CC IsoId=Q9VUE5-2; Sequence=VSP\_050479;  
 CC Note=No experimental confirmation available;  
 CC Name=C;  
 CC IsoId=Q9VUE5-3; Sequence=VSP\_050480;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Specifically expressed at the tip of the  
 CC furrow in cellularizing blastoderms. CLIP-190 and jar are  
 CC coexpressed at several times in development and in a number of  
 CC tissues, including embryonic axonal neuron processes and posterior  
 CC pole.  
 CC -1- SIMILARITY: Contains 2 CAP-Gly domains.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; AF041382; AAB96783.1; -  
 CC EMBL; AB003655; AAF53604.1; -  
 CC EMBL; AB003655; AAF53605.2; -  
 CC EMBL; AB003655; AAN10987.1; -  
 CC EMBL; AY118896; AAM50756.1; -  
 CC Flybase; FBgn020503; CLIP-190.  
 CC GO; GO:0005938; C:cell cortex; IDA.  
 CC GO; GO:0005794; C:Golgi apparatus; IDA.  
 CC GO; GO:0005875; C:microtubule associated complex; IDA.  
 CC GO; GO:0003779; F:actin binding; IDA.  
 CC GO; GO:0008017; F:microtubule binding; IDA.  
 CC GO; GO:0005515; F:protein binding; IPI.  
 CC GO; GO:0007349; P:cellularization; IMP.  
 CC InterPro; IPR000936; CAP-Gly.  
 CC Pfam; PF01302; CAP GLY; 2.  
 CC PROSITE; PS00845; CAP GLY 1; FALSE\_NEG.  
 CC PROSITE; PS02445; CAP\_GLY\_2; 2.

Cytoskeleton; Golgi stack; Microtubule; Actin-binding; Coiled coil;  
Repeat; Alternative splicing. CAP-GLY 1.  
KW DOMAIN 143 185  
FT DOMAIN 260 302  
FT DOMAIN 378 468  
FT DOMAIN 484 660  
FT DOMAIN 488 1452  
FT DOMAIN 667 916  
FT DOMAIN 926 981  
FT DOMAIN 1001 1121  
FT DOMAIN 1158 1549  
FT DOMAIN 1200 1461  
FT DOMAIN 1565 1600  
FT VARSPLIC 348 348  
FT VARSPLIC 1 109

MSDDTSASGSGSAPSPSVTADPPGATASKLPGPIRSNIP  
TPATSGTIPQSKADSPSGSVSKGRPCNCHTTPK  
SGPPREATSRSDNLSINSAYT -> MSRESDNLS  
SINRAYTLYQETVRRTPRSLSPTPDWFRSPARRLSKSE  
AGRSAYDYLYEATGRSS (in isoform  
C).  
/FTid=VSP\_050480.  
S -> N (IN REP. 1).  
D -> G (IN REP. 1).  
K -> Q (IN REP. 1).  
E -> A (IN REP. 1).  
T -> S (IN REP. 1).  
S -> I (IN REP. 1).  
N -> Q (IN REP. 1).  
M -> K (IN REP. 1).  
F -> L (IN REP. 1).  
Q -> E (IN REP. 1).  
HLL -> QLO (IN REP. 1).  
C -> Y (IN REP. 1).  
189063 MW; D6F7916A9532F16 CRG64;  
9.1%; Score 253.5; DB 1; Length 1690;  
Best Local Similarity 23.3%; Pred. No. 0.0048;  
Matches 127; Conservative 102; Mismatches 180; Indels 135; Gaps 23;

49 SQAPRKDEGAQRTAQSGALDVSELSRQLEDILSTVCYDNNQGGPGEDGAQCEPAEPE 108  
352 SSTVPKPIATPKSQFS--MDLREKQKHQVKKLVVERDLDR-----E 392

109 DAESRTYVARGSPPEPTPVYMGKSPKSGDPNTEIRQS--DEV---GDRDHRPQ-EKK 163  
393 DAQNALQLOQNLKARIVLESALDNRKTEELQCSIDEAQPCGDELNAQSQVYKE 452

164 KAKGLGKEITLMQ---TLNLTSTP-----EKLALCKK-----YAEEL 200  
453 KIHDLSEKITKLVSATPSLQSLPDLPSDDGALQEEIAKLQEKMTIQKEVESRIAEOL 512

201 BEHNSQOMKLLQKX---OSQVQEKDHLRGHNSKAVLARSKLSLCRELORHNSLK 256  
513 EEEQLRENVKYLNEQIATLOSELVSKDEAL-----EKFSLSECGIENLRLE---LLK 564

257 EGVQVRAPEE-----EKKVY-----TSHFOVTIANDI 284  
565 EENEKQAEQAQAEFTPKLAESVYVRLSELQNLKATSDSESERVKNKTDECEILQTEV 624

285 QLOMEQHNRNSKL---RQNMELAE-----RLKGLQEYVELREHIDKVFKEHL 332  
625 RMRDEQRELNQQLDEVTQLNVQKASDALLMLRQK--EGTEEKSTULEKT----- 676

333 QQQLVDAKLQAOQEMLKEABERHOREKDFLLKEAVESQRMCELMKQ-----QETHLK 384  
677 EKELVQSK-EQAATXNDKEQLKQISD--LKQALAEQELKLVREMTENAINQIQLEKESIE 733

385 QQLALYTEKEEFPQNTLSKSEVTTTFQ-----EKEVTKIKLKEETMYRWRWE 437  
734 QQLALKQNELEDFOKQKQSEVHVHQLQETKQNTQKDFELVSGESLKKLQQLQEQTLGHE 793

Query Match  
Best Local Similarity 9.1%; Score 253.5; DB 1; Length 1690;  
Matches 127; Conservative 102; Mismatches 180; Indels 135; Gaps 23;

49 SQAPRKDEGAQRTAQSGALDVSELSRQLEDILSTVCYDNNQGGPGEDGAQCEPAEPE 108  
352 SSTVPKPIATPKSQFS--MDLREKQKHQVKKLVVERDLDR-----E 392

109 DAESRTYVARGSPPEPTPVYMGKSPKSGDPNTEIRQS--DEV---GDRDHRPQ-EKK 163  
393 DAQNALQLOQNLKARIVLESALDNRKTEELQCSIDEAQPCGDELNAQSQVYKE 452

164 KAKGLGKEITLMQ---TLNLTSTP-----EKLALCKK-----YAEEL 200  
453 KIHDLSEKITKLVSATPSLQSLPDLPSDDGALQEEIAKLQEKMTIQKEVESRIAEOL 512

201 BEHNSQOMKLLQKX---OSQVQEKDHLRGHNSKAVLARSKLSLCRELORHNSLK 256  
513 EEEQLRENVKYLNEQIATLOSELVSKDEAL-----EKFSLSECGIENLRLE---LLK 564

257 EGVQVRAPEE-----EKKVY-----TSHFOVTIANDI 284  
565 EENEKQAEQAQAEFTPKLAESVYVRLSELQNLKATSDSESERVKNKTDECEILQTEV 624

285 QLOMEQHNRNSKL---RQNMELAE-----RLKGLQEYVELREHIDKVFKEHL 332  
625 RMRDEQRELNQQLDEVTQLNVQKASDALLMLRQK--EGTEEKSTULEKT----- 676

333 QQQLVDAKLQAOQEMLKEABERHOREKDFLLKEAVESQRMCELMKQ-----QETHLK 384  
677 EKELVQSK-EQAATXNDKEQLKQISD--LKQALAEQELKLVREMTENAINQIQLEKESIE 733

385 QQLALYTEKEEFPQNTLSKSEVTTTFQ-----EKEVTKIKLKEETMYRWRWE 437  
734 QQLALKQNELEDFOKQKQSEVHVHQLQETKQNTQKDFELVSGESLKKLQQLQEQTLGHE 793

438 SSNNKALLEMAEKTIV-----RDKLEGLQVQIKORLEKLCRALQTERNDINRQVQLSAGGQ 493  
794 KLAQALEELKKEETIKKEQEQLQQLQSKSAESALKVQVQLEQLQ--QAAASGEE 851

494 GSLT 497  
852 GSKT 855

RESULT 12  
MYH4 RABIT  
ID MYH4 RABIT STANDARD; PRT; 1938 AA.  
AC Q28641;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Myosin heavy chain, skeletal muscle, juvenile.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;  
RA Maeda K., Hostinova E., Roesch-Kleinlauf A., Schuster H., Gasperik J.,  
RA Wittinghofer A.;  
RT "Isolation, sequencing of myosin heavy chain cDNA from rabbit  
skeletal muscle and a novel cosynthesis of S-1 fragment with the  
essential and regulatory light chains.";  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
CC -|- FUNCTION: Muscle contraction.  
CC -|- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2  
heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)  
and 2 regulatory light chain subunits (MLC-2).  
CC -|- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
CC -|- DOMAIN: The rodlike tail sequence is highly repetitive, showing  
cycles of a 28-residue repeat pattern composed of 4 heptapeptides,  
characteristic for alpha-helical coiled coils.  
CC -|- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light  
meromyosin (LMM) and 1 heavy meromyosin (HMM). It can later be  
split further into 2 globular subfragments (S1) and 1 rod-shaped  
subfragment (S2).  
CC -|- SIMILARITY: Contains 1 myosin-like globular head domain.  
CC -|- SIMILARITY: Contains 1 IQ domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U32574; AAA74199.1; -;  
PIR; A59293; A59293.  
HSP; P13538; 2MYS.  
InterPro; IPR000048; IQ\_region.  
InterPro; IPR001609; myosin\_head.  
InterPro; IPR004009; Myosin\_N.  
InterPro; IPR002928; Myosin\_tail.  
Pfam; PF00612; IQ; 2.  
Pfam; PF00063; myosin\_head; 1.  
Pfam; PF02736; Myosin\_N; 1.  
Pfam; PF01576; Myosin\_tail; 1.  
PRINTS; P00193; MYOSINHEAVY.  
ProDom; PD000355; myosin\_head; 1.  
SMART; SM00015; IQ; 1.  
SMART; SM00242; MYC; 1.  
PROSITE; PS50056; IQ; 1.  
Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
Calmodulin-binding; ATP-binding; Methylation; Multi-gene family.  
DOMAIN 1 783 MYOSIN HEAD-LIKE.

```
FT DOMAIN 784 813 IQ.
FT DOMAIN 842 1938
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 658 680 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 35 35 METHYLATION (MONO-) (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).
SQ SEQUENCE 1938 AA; 223064 MW; D8A8A2EC5B182626 CRC64;

Query Match
Best Local Similarity 8.9%; Score 249; DB 1; Length 1938;
Matches 113; Conservative 80; Mismatches 179; Indels 160; Gaps 16;

QY 49 SQAKPKPGQAQTAQSGALDVEELSRQLEDILSTVCYVNNQGGPGDGAQGEPAPE 108
DB 1091 SNLQSKIDQALQW--LQKIKELQATEEL-----EEIEIERASRA 1133
QY 109 DAKSRVTYVANGBEPTPVVNGEKEPSKDPNTEIRQS--DEVDRRHRRPQEKKAAG 167
DB 1134 KAEKQSDLSR-----ELESERLEERAGGATSAGIEMNKREA 1172
QY 168 LGKETILMOTLNTLSTPEKLAALCKK-----YAELEERNQKQKLLQKXOSOLVQ 222
DB 1173 ---EFQKMRDLERATLQHEATAATLRKKHADSVAEELGCEQIDNLRVQKLEKEKSELKM 1229
QY 223 EKDLHGEHKAVALARSKLESICREL-----QRNRSLEKRGVQAR-- 264
DB 1230 EIDDLASNETVSKAGNLEKWCRTLEDQVSELTKKEEHRLLINDLSAQARLOTESGE 1289
QY 265 -----BEEKREKVTSH-FQVTLNDIQLQMEQ 290
DB 1290 FSRQLDEKDSLVLSQSRGQAFTQIEELKQLESEETKAKSALAHQASRHDCLLREQ 1349
QY 291 HNER-----NSKL-----QENNELAERLKKLLEQVLEEEHDX 325
DB 1350 YEEQEAKEALQRAWKANSEVAMQRTKYETATLQRTTELEBAKKLAQLQDAEEHV-- 1407
QY 326 VFKHDLQOQLVDKLAQAQMLKEAEERHOREKDFLLKEAVESQRCMLAKQOETHLKQ 385
DB 1408 -----EAVNAKAS-----LEKTKQLQNEVEDLMIDVETNAACAALDKQRPDX 1454
QY 386 QIALYTEKPEFPQNTLSKSSVFYTFKQEMEMTKKIKLEKETITVRSRWESNKALLE 445
DB 1455 ILAEWKHKYETH-----AELEASQKESRSLSTEVFKVKNAYEESLDQLET 1500
QY 446 MAEKTVDKLEGL--QV-----KIQLEKLCALQTERNDLNKRVQDLQA 490
DB 1501 LKREKNEQQISDLTEQIAGGKRIHLEKVKQVEQEKSELQALEEABA 1552

RESULT 13
RRB1 MOUSE
ID_RRB1 MOUSE STANDARD; PRT; 1605 AA.
AC Q99PL5; Q99PK5; Q99PK6; Q99PK7; Q99PK8; Q99PK9; Q99PL0; Q99PL1;
AC Q99PL2; Q99PL3; Q99PL4; Q9CS20;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosome-binding protein 1 (Ribosome receptor protein) (mRrp).
GN RRPp1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS RRP61/P180; RRP47; RRP41; RRP16.8;
RP RRP15A; RRP15B; RRP10; RRP5.4; RRP2; RRP1.8 AND RRP0), AND TISSUE
RP SPECIFICITY.
RC STRAIN=IC3; TISSUE=Embryo;
RX MEDLINE=21104429; PubMed=11167022;
RA Kim Y.-J., Lee M.-C., Kim S.-J., Chun J.-Y.;
```

```
*Identification and characterization of multiple isoforms of a mouse
ribosome receptor.*;
Gene 261:337-344 (2000).
[2]
SEQUENCE FROM N.A. (ISOFORM 2).
TISSUE=Kidney;
MEDLINE=22388257; PubMed=12477932;
Strauber R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
Klauniger R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallya S.J.,
Rosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Vallaloun D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Whiting J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
*Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.*;
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[3]
SEQUENCE OF 736-1605 FROM N.A. (ISOFORM 3).
STRAIN=C57BL/6J; TISSUE=Embryo;
MEDLINE=21085660; PubMed=11217851;
Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
Kadoh K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Blake J., Offord T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzaresi J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
Hayashizaki Y.;
*Functional annotation of a full-length mouse cDNA collection.*;
Nature 409:685-690 (2001).
-!- FUNCTION: Acts as a ribosome receptor and mediates interaction
between the ribosome and the endoplasmic reticulum membrane (by
similarity).
-!- SUBCELLULAR LOCATION: Type III membrane protein. Endoplasmic
reticulum (by similarity).
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=13;
Comment=Additional isoforms seem to exist. Experimental
confirmation may be lacking for some isoforms;
Name=3;
IsoId=Q99PL5-1; Sequence=Displayed;
Name=RRP61; Synonym=ms-p180;
IsoId=Q99PL5-2; Sequence=VSP_003963, VSP_003964;
Name=RRP47;
IsoId=Q99PL5-3; Sequence=VSP_003959, VSP_003963, VSP_003964;
Name=RRP41;
IsoId=Q99PL5-4; Sequence=VSP_003961, VSP_003963, VSP_003964;
Name=RRP16.8;
IsoId=Q99PL5-5; Sequence=VSP_003954, VSP_003962, VSP_003963,
VSP_003964;
Name=RRP15a;
IsoId=Q99PL5-6; Sequence=VSP_003958, VSP_003963, VSP_003964;
```





GN PLECTIN  
OS Cricetulus griseus (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetulus.  
OX NCBI\_TaxID=10029;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20334248; PubMed=10873583;  
RA Clubb B.H.; Chou Y.-H.; Herrmann H.; Svitek T.M.; Borisy G.G.;  
RT "The 300-kDa intermediate filament-associated protein (IFAP300) is a  
RT hamster plectin ortholog.";  
RL Biochem. Biophys. Res. Commun. 273:183-187(2000).  
RN [2]  
RP PHOSPHORYLATION.  
RX MEDLINE=96215219; PubMed=8626512;  
RA Malec N.; Foisner R.; Stadler C.; Wiche G.;  
RT "Identification of plectin as a substrate of p34cdc2 kinase and  
RT mapping of a single phosphorylation site.";  
RL J. Biol. Chem. 271:8203-8208(1996).  
CC -!- FUNCTION: Interlinks intermediate filaments with microtubules and  
CC microfilaments and anchors intermediate filaments to desmosomes or  
CC hemidesmosomes. May be involved not only in the crosslinking and  
CC stabilization of cytoskeletal intermediate filaments network, but  
CC also in the regulation of their dynamics.  
CC -!- SUBUNIT: Homodimer or homotetramer.  
CC -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with  
CC vimentin, desmin, GFAP, cytokeratins, lamin B; whereas both the N-  
CC and the C-terminus can bind integrin beta-4.  
CC -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM  
CC INTERMEDIATE FILAMENTS DURING MITOSIS.  
CC -!- SIMILARITY: Contains 1 actin-binding domain.  
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC -!- SIMILARITY: Contains 32 plectin repeats.  
CC -!- SIMILARITY: Contains 4 spectrin repeats.  
CC -!- SIMILARITY: Belongs to the plakin or cycolinker family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; AF260753; AAF70372.1; -  
CC HSP; Q01082; 1BKX.  
CC InterPro; IPR001589; Actbind actinin.  
CC InterPro; IPR001715; Calponin-like.  
CC InterPro; IPR001101; Plectin repeat.  
CC InterPro; IPR002017; Spectrin.  
CC Pfam; PF00307; CH; 2.  
CC Pfam; PF00681; Plectin; 20.  
CC SMART; SMO0033; CH; 2.  
CC SMART; SMO0250; PLEC; 34.  
CC PROSITE; PS00019; ACTININ\_1; PARTIAL.  
CC PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
CC PROSITE; PS50021; CH; 2.  
CC Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;  
CC Phosphorylation.  
CC NON TER 1  
FT DOMAIN 1  
FT 1259  
FT 1260  
FT 1261  
FT 1262  
FT 1263  
FT 1264  
FT 1265  
FT 1266  
FT 1267  
FT 1268  
FT 1269  
FT 1270  
FT 1271  
FT 1272  
FT 1273  
FT 1274  
FT 1275  
FT 1276  
FT 1277  
FT 1278  
FT 1279  
FT 1280  
FT 1281  
FT 1282  
FT 1283  
FT 1284  
FT 1285  
FT 1286  
FT 1287  
FT 1288  
FT 1289  
FT 1290  
FT 1291  
FT 1292  
FT 1293  
FT 1294  
FT 1295  
FT 1296  
FT 1297  
FT 1298  
FT 1299  
FT 1300  
FT 1301  
FT 1302  
FT 1303  
FT 1304  
FT 1305  
FT 1306  
FT 1307  
FT 1308  
FT 1309  
FT 1310  
FT 1311  
FT 1312  
FT 1313  
FT 1314  
FT 1315  
FT 1316  
FT 1317  
FT 1318  
FT 1319  
FT 1320  
FT 1321  
FT 1322  
FT 1323  
FT 1324  
FT 1325  
FT 1326  
FT 1327  
FT 1328  
FT 1329  
FT 1330  
FT 1331  
FT 1332  
FT 1333  
FT 1334  
FT 1335  
FT 1336  
FT 1337  
FT 1338  
FT 1339  
FT 1340  
FT 1341  
FT 1342  
FT 1343  
FT 1344  
FT 1345  
FT 1346  
FT 1347  
FT 1348  
FT 1349  
FT 1350  
FT 1351  
FT 1352  
FT 1353  
FT 1354  
FT 1355  
FT 1356  
FT 1357  
FT 1358  
FT 1359  
FT 1360  
FT 1361  
FT 1362  
FT 1363  
FT 1364  
FT 1365  
FT 1366  
FT 1367  
FT 1368  
FT 1369  
FT 1370  
FT 1371  
FT 1372  
FT 1373  
FT 1374  
FT 1375  
FT 1376  
FT 1377  
FT 1378  
FT 1379  
FT 1380  
FT 1381  
FT 1382  
FT 1383  
FT 1384  
FT 1385  
FT 1386  
FT 1387  
FT 1388  
FT 1389  
FT 1390  
FT 1391  
FT 1392  
FT 1393  
FT 1394  
FT 1395  
FT 1396  
FT 1397  
FT 1398  
FT 1399  
FT 1400  
FT 1401  
FT 1402  
FT 1403  
FT 1404  
FT 1405  
FT 1406  
FT 1407  
FT 1408  
FT 1409  
FT 1410  
FT 1411  
FT 1412  
FT 1413  
FT 1414  
FT 1415  
FT 1416  
FT 1417  
FT 1418  
FT 1419  
FT 1420  
FT 1421  
FT 1422  
FT 1423  
FT 1424  
FT 1425  
FT 1426  
FT 1427  
FT 1428  
FT 1429  
FT 1430  
FT 1431  
FT 1432  
FT 1433  
FT 1434  
FT 1435  
FT 1436  
FT 1437  
FT 1438  
FT 1439  
FT 1440  
FT 1441  
FT 1442  
FT 1443  
FT 1444  
FT 1445  
FT 1446  
FT 1447  
FT 1448  
FT 1449  
FT 1450  
FT 1451  
FT 1452  
FT 1453  
FT 1454  
FT 1455  
FT 1456  
FT 1457  
FT 1458  
FT 1459  
FT 1460  
FT 1461  
FT 1462  
FT 1463  
FT 1464  
FT 1465  
FT 1466  
FT 1467  
FT 1468  
FT 1469  
FT 1470  
FT 1471  
FT 1472  
FT 1473  
FT 1474  
FT 1475  
FT 1476  
FT 1477  
FT 1478  
FT 1479  
FT 1480  
FT 1481  
FT 1482  
FT 1483  
FT 1484  
FT 1485  
FT 1486  
FT 1487  
FT 1488  
FT 1489  
FT 1490  
FT 1491  
FT 1492  
FT 1493  
FT 1494  
FT 1495  
FT 1496  
FT 1497  
FT 1498  
FT 1499  
FT 1500  
FT 1501  
FT 1502  
FT 1503  
FT 1504  
FT 1505  
FT 1506  
FT 1507  
FT 1508  
FT 1509  
FT 1510  
FT 1511  
FT 1512  
FT 1513  
FT 1514  
FT 1515  
FT 1516  
FT 1517  
FT 1518  
FT 1519  
FT 1520  
FT 1521  
FT 1522  
FT 1523  
FT 1524  
FT 1525  
FT 1526  
FT 1527  
FT 1528  
FT 1529  
FT 1530  
FT 1531  
FT 1532  
FT 1533  
FT 1534  
FT 1535  
FT 1536  
FT 1537  
FT 1538  
FT 1539  
FT 1540  
FT 1541  
FT 1542  
FT 1543  
FT 1544  
FT 1545  
FT 1546  
FT 1547  
FT 1548  
FT 1549  
FT 1550  
FT 1551  
FT 1552  
FT 1553  
FT 1554  
FT 1555  
FT 1556  
FT 1557  
FT 1558  
FT 1559  
FT 1560  
FT 1561  
FT 1562  
FT 1563  
FT 1564  
FT 1565  
FT 1566  
FT 1567  
FT 1568  
FT 1569  
FT 1570  
FT 1571  
FT 1572  
FT 1573  
FT 1574  
FT 1575  
FT 1576  
FT 1577  
FT 1578  
FT 1579  
FT 1580  
FT 1581  
FT 1582  
FT 1583  
FT 1584  
FT 1585  
FT 1586  
FT 1587  
FT 1588  
FT 1589  
FT 1590  
FT 1591  
FT 1592  
FT 1593  
FT 1594  
FT 1595  
FT 1596  
FT 1597  
FT 1598  
FT 1599  
FT 1600  
FT 1601  
FT 1602  
FT 1603  
FT 1604  
FT 1605  
FT 1606  
FT 1607  
FT 1608  
FT 1609  
FT 1610  
FT 1611  
FT 1612  
FT 1613  
FT 1614  
FT 1615  
FT 1616  
FT 1617  
FT 1618  
FT 1619  
FT 1620  
FT 1621  
FT 1622  
FT 1623  
FT 1624  
FT 1625  
FT 1626  
FT 1627  
FT 1628  
FT 1629  
FT 1630  
FT 1631  
FT 1632  
FT 1633  
FT 1634  
FT 1635  
FT 1636  
FT 1637  
FT 1638  
FT 1639  
FT 1640  
FT 1641  
FT 1642  
FT 1643  
FT 1644  
FT 1645  
FT 1646  
FT 1647  
FT 1648  
FT 1649  
FT 1650  
FT 1651  
FT 1652  
FT 1653  
FT 1654  
FT 1655  
FT 1656  
FT 1657  
FT 1658  
FT 1659  
FT 1660  
FT 1661  
FT 1662  
FT 1663  
FT 1664  
FT 1665  
FT 1666  
FT 1667  
FT 1668  
FT 1669  
FT 1670  
FT 1671  
FT 1672  
FT 1673  
FT 1674  
FT 1675  
FT 1676  
FT 1677  
FT 1678  
FT 1679  
FT 1680  
FT 1681  
FT 1682  
FT 1683  
FT 1684  
FT 1685  
FT 1686  
FT 1687  
FT 1688  
FT 1689  
FT 1690  
FT 1691  
FT 1692  
FT 1693  
FT 1694  
FT 1695  
FT 1696  
FT 1697  
FT 1698  
FT 1699  
FT 1700  
FT 1701  
FT 1702  
FT 1703  
FT 1704  
FT 1705  
FT 1706  
FT 1707  
FT 1708  
FT 1709  
FT 1710  
FT 1711  
FT 1712  
FT 1713  
FT 1714  
FT 1715  
FT 1716  
FT 1717  
FT 1718  
FT 1719  
FT 1720  
FT 1721  
FT 1722  
FT 1723  
FT 1724  
FT 1725  
FT 1726  
FT 1727  
FT 1728  
FT 1729  
FT 1730  
FT 1731  
FT 1732  
FT 1733  
FT 1734  
FT 1735  
FT 1736  
FT 1737  
FT 1738  
FT 1739  
FT 1740  
FT 1741  
FT 1742  
FT 1743  
FT 1744  
FT 1745  
FT 1746  
FT 1747  
FT 1748  
FT 1749  
FT 1750  
FT 1751  
FT 1752  
FT 1753  
FT 1754  
FT 1755  
FT 1756  
FT 1757  
FT 1758  
FT 1759  
FT 1760  
FT 1761  
FT 1762  
FT 1763  
FT 1764  
FT 1765  
FT 1766  
FT 1767  
FT 1768  
FT 1769  
FT 1770  
FT 1771  
FT 1772  
FT 1773  
FT 1774  
FT 1775  
FT 1776  
FT 1777  
FT 1778  
FT 1779  
FT 1780  
FT 1781  
FT 1782  
FT 1783  
FT 1784  
FT 1785  
FT 1786  
FT 1787  
FT 1788  
FT 1789  
FT 1790  
FT 1791  
FT 1792  
FT 1793  
FT 1794  
FT 1795  
FT 1796  
FT 1797  
FT 1798  
FT 1799  
FT 1800  
FT 1801  
FT 1802  
FT 1803  
FT 1804  
FT 1805  
FT 1806  
FT 1807  
FT 1808  
FT 1809  
FT 1810  
FT 1811  
FT 1812  
FT 1813  
FT 1814  
FT 1815  
FT 1816  
FT 1817  
FT 1818  
FT 1819  
FT 1820  
FT 1821  
FT 1822  
FT 1823  
FT 1824  
FT 1825  
FT 1826  
FT 1827  
FT 1828  
FT 1829  
FT 1830  
FT 1831  
FT 1832  
FT 1833  
FT 1834  
FT 1835  
FT 1836  
FT 1837  
FT 1838  
FT 1839  
FT 1840  
FT 1841  
FT 1842  
FT 1843  
FT 1844  
FT 1845  
FT 1846  
FT 1847  
FT 1848  
FT 1849  
FT 1850  
FT 1851  
FT 1852  
FT 1853  
FT 1854  
FT 1855  
FT 1856  
FT 1857  
FT 1858  
FT 1859  
FT 1860  
FT 1861  
FT 1862  
FT 1863  
FT 1864  
FT 1865  
FT 1866  
FT 1867  
FT 1868  
FT 1869  
FT 1870  
FT 1871  
FT 1872  
FT 1873  
FT 1874  
FT 1875  
FT 1876  
FT 1877  
FT 1878  
FT 1879  
FT 1880  
FT 1881  
FT 1882  
FT 1883  
FT 1884  
FT 1885  
FT 1886  
FT 1887  
FT 1888  
FT 1889  
FT 1890  
FT 1891  
FT 1892  
FT 1893  
FT 1894  
FT 1895  
FT 1896  
FT 1897  
FT 1898  
FT 1899  
FT 1900  
FT 1901  
FT 1902  
FT 1903  
FT 1904  
FT 1905  
FT 1906  
FT 1907  
FT 1908  
FT 1909  
FT 1910  
FT 1911  
FT 1912  
FT 1913  
FT 1914  
FT 1915  
FT 1916  
FT 1917  
FT 1918  
FT 1919  
FT 1920  
FT 1921  
FT 1922  
FT 1923  
FT 1924  
FT 1925  
FT 1926  
FT 1927  
FT 1928  
FT 1929  
FT 1930  
FT 1931  
FT 1932  
FT 1933  
FT 1934  
FT 1935  
FT 1936  
FT 1937  
FT 1938  
FT 1939  
FT 1940  
FT 1941  
FT 1942  
FT 1943  
FT 1944  
FT 1945  
FT 1946  
FT 1947  
FT 1948  
FT 1949  
FT 1950  
FT 1951  
FT 1952  
FT 1953  
FT 1954  
FT 1955  
FT 1956  
FT 1957  
FT 1958  
FT 1959  
FT 1960  
FT 1961  
FT 1962  
FT 1963  
FT 1964  
FT 1965  
FT 1966  
FT 1967  
FT 1968  
FT 1969  
FT 1970  
FT 1971  
FT 1972  
FT 1973  
FT 1974  
FT 1975  
FT 1976  
FT 1977  
FT 1978  
FT 1979  
FT 1980  
FT 1981  
FT 1982  
FT 1983  
FT 1984  
FT 1985  
FT 1986  
FT 1987  
FT 1988  
FT 1989  
FT 1990  
FT 1991  
FT 1992  
FT 1993  
FT 1994  
FT 1995  
FT 1996  
FT 1997  
FT 1998  
FT 1999  
FT 2000  
FT 2001  
FT 2002  
FT 2003  
FT 2004  
FT 2005  
FT 2006  
FT 2007  
FT 2008  
FT 2009  
FT 2010  
FT 2011  
FT 2012  
FT 2013  
FT 2014  
FT 2015  
FT 2016  
FT 2017  
FT 2018  
FT 2019  
FT 2020  
FT 2021  
FT 2022  
FT 2023  
FT 2024  
FT 2025  
FT 2026  
FT 2027  
FT 2028  
FT 2029  
FT 2030  
FT 2031  
FT 2032  
FT 2033  
FT 2034  
FT 2035  
FT 2036  
FT 2037  
FT 2038  
FT 2039  
FT 2040  
FT 2041  
FT 2042  
FT 2043  
FT 2044  
FT 2045  
FT 2046  
FT 2047  
FT 2048  
FT 2049  
FT 2050  
FT 2051  
FT 2052  
FT 2053  
FT 2054  
FT 2055  
FT 2056  
FT 2057  
FT 2058  
FT 2059  
FT 2060  
FT 2061  
FT 2062  
FT 2063  
FT 2064  
FT 2065  
FT 2066  
FT 2067  
FT 2068  
FT 2069  
FT 2070  
FT 2071  
FT 2072  
FT 2073  
FT 2074  
FT 2075  
FT 2076  
FT 2077  
FT 2078  
FT 2079  
FT 2080  
FT 2081  
FT 2082  
FT 2083  
FT 2084  
FT 2085  
FT 2086  
FT 2087  
FT 2088  
FT 2089  
FT 2090  
FT 2091  
FT 2092  
FT 2093  
FT 2094  
FT 2095  
FT 2096  
FT 2097  
FT 2098  
FT 2099  
FT 2100  
FT 2101  
FT 2102  
FT 2103  
FT 2104  
FT 2105  
FT 2106  
FT 2107  
FT 2108  
FT 2109  
FT 2110  
FT 2111  
FT 2112  
FT 2113  
FT 2114  
FT 2115  
FT 2116  
FT 2117  
FT 2118  
FT 2119  
FT 2120  
FT 2121  
FT 2122  
FT 2123  
FT 2124  
FT 2125  
FT 2126  
FT 2127  
FT 2128  
FT 2129  
FT 2130  
FT 2131  
FT 2132  
FT 2133  
FT 2134  
FT 2135  
FT 2136  
FT 2137  
FT 2138  
FT 2139  
FT 2140  
FT 2141  
FT 2142  
FT 2143  
FT 2144  
FT 2145  
FT 2146  
FT 2147  
FT 2148  
FT 2149  
FT 2150  
FT 2151  
FT 2152  
FT 2153  
FT 2154  
FT 2155  
FT 2156  
FT 2157  
FT 2158  
FT 2159  
FT 2160  
FT 2161  
FT 2162  
FT 2163  
FT 2164  
FT 2165  
FT 2166  
FT 2167  
FT 2168  
FT 2169  
FT 2170  
FT 2171  
FT 2172  
FT 2173  
FT 2174  
FT 2175  
FT 2176  
FT 2177  
FT 2178  
FT 2179  
FT 2180  
FT 2181  
FT 2182  
FT 2183  
FT 2184  
FT 2185  
FT 2186  
FT 2187  
FT 2188  
FT 2189  
FT 2190  
FT 2191  
FT 2192  
FT 2193  
FT 2194  
FT 2195  
FT 2196  
FT 2197  
FT 2198  
FT 2199  
FT 2200  
FT 2201  
FT 2202  
FT 2203  
FT 2204  
FT 2205  
FT 2206  
FT 2207  
FT 2208  
FT 2209  
FT 2210  
FT 2211  
FT 2212  
FT 2213  
FT 2214  
FT 2215  
FT 2216  
FT 2217  
FT 2218  
FT 2219  
FT 2220  
FT 2221  
FT 2222  
FT 2223  
FT 2224  
FT 2225  
FT 2226  
FT 2227  
FT 2228  
FT 2229  
FT 2230  
FT 2231  
FT 2232  
FT 2233  
FT 2234  
FT 2235  
FT 2236  
FT 2237  
FT 2238  
FT 2239  
FT 2240  
FT 2241  
FT 2242  
FT 2243  
FT 2244  
FT 2245  
FT 2246  
FT 2247  
FT 2248  
FT 2249  
FT 2250  
FT 2251  
FT 2252  
FT 2253  
FT 2254  
FT 2255  
FT 2256  
FT 2257  
FT 2258  
FT 2259  
FT 2260  
FT 2261  
FT 2262  
FT 2263  
FT 2264  
FT 2265  
FT 2266  
FT 2267  
FT 2268  
FT 2269  
FT 2270  
FT 2271  
FT 2272  
FT 2273  
FT 2274  
FT 2275  
FT 2276  
FT 2277  
FT 2278  
FT 2279  
FT 2280  
FT 2281  
FT 2282  
FT 2283  
FT 2284  
FT 2285  
FT 2286  
FT 2287  
FT 2288  
FT 2289  
FT 2290  
FT 2291  
FT 2292  
FT 2293  
FT 2294  
FT 2295  
FT 2296  
FT 2297  
FT 2298  
FT 2299  
FT 2300  
FT 2301  
FT 2302  
FT 2303  
FT 2304  
FT 2305  
FT 2306  
FT 2307  
FT 2308  
FT 2309  
FT 2310  
FT 2311  
FT 2312  
FT 2313  
FT 2314  
FT 2315  
FT 2316  
FT 2317  
FT 2318  
FT 2319  
FT 2320  
FT 2321  
FT 2322  
FT 2323  
FT 2324  
FT 2325  
FT 2326  
FT 2327  
FT 2328  
FT 2329  
FT 2330  
FT 2331  
FT 2332  
FT 2333  
FT 2334  
FT 2335  
FT 2336  
FT 2337  
FT 2338  
FT 2339  
FT 2340  
FT 2341  
FT 2342  
FT 2343  
FT 2344  
FT 2345  
FT 2346  
FT 2347  
FT 2348  
FT 2349  
FT 2350  
FT 2351  
FT 2352  
FT 2353  
FT 2354  
FT 2355  
FT 2356  
FT 2357  
FT 2358  
FT 2359  
FT 2360  
FT 2361  
FT 2362  
FT 2363  
FT 2364  
FT 2365  
FT 2366  
FT 2367  
FT 2368  
FT 2369  
FT 2370  
FT 2371  
FT 2372  
FT 2373  
FT 2374  
FT 2375  
FT 2376  
FT 2377  
FT 2378  
FT 2379  
FT 2380  
FT 2381  
FT 2382  
FT 2383  
FT 2384  
FT 2385  
FT 2386  
FT 2387  
FT 2388  
FT 2389  
FT 2390  
FT 2391  
FT 2392  
FT 2393  
FT 2394  
FT 2395  
FT 2396  
FT 2397  
FT 2398  
FT 2399  
FT 2400  
FT 2401  
FT 2402  
FT 2403  
FT 2404  
FT 2405  
FT 2406  
FT 2407  
FT 2408  
FT 2409  
FT 2410  
FT 2411  
FT 2412  
FT 2413  
FT 2414  
FT 2415  
FT 2416  
FT 2417  
FT 2418  
FT 2419  
FT 2420  
FT 2421  
FT 2422  
FT 2423  
FT 2424  
FT 2425  
FT 2426  
FT 2427  
FT 2428  
FT 2429  
FT 2430  
FT 2431  
FT 2432  
FT 2433  
FT 2434  
FT 2435  
FT 2436  
FT 2437  
FT 2438  
FT 2439  
FT 2440  
FT 2441  
FT 2442  
FT 2443  
FT 2444  
FT 2445  
FT 2446  
FT 2447  
FT 2448  
FT 2449  
FT 2450  
FT 2451  
FT 2452  
FT 2453  
FT 2454  
FT 2455  
FT 2456  
FT 2457  
FT 2458  
FT 2459  
FT 2460  
FT 2461  
FT 2462  
FT 2463  
FT 2464  
FT 2465  
FT 2466  
FT 2467  
FT 2468  
FT 2469  
FT 2470  
FT 2471  
FT 2472  
FT 2473  
FT 2474  
FT 2475  
FT 2476  
FT 2477  
FT 2478  
FT 2479  
FT 2480  
FT 2481  
FT 2482  
FT 2483  
FT 2484  
FT 2485  
FT 2486  
FT 2487  
FT 2488  
FT 2489  
FT 2490  
FT 2491  
FT 2492  
FT 2493  
FT 2494  
FT 2495  
FT 2496  
FT 2497  
FT 2498  
FT 2499  
FT 2500  
FT 2501  
FT 2502  
FT 2503  
FT 2504  
FT 2505  
FT 2506  
FT 2507  
FT 2508  
FT 2509  
FT 2510  
FT 2511  
FT 2512  
FT 2513  
FT 2514  
FT 2515  
FT 2516  
FT 2517  
FT 2518  
FT 2519  
FT 2520  
FT 2521  
FT 2522  
FT 2523  
FT 2524  
FT 2525  
FT 2526  
FT 2527  
FT 2528  
FT 2529  
FT 2530  
FT 2531  
FT 2532  
FT 2533  
FT 2534  
FT 2535  
FT 2536  
FT 2537  
FT 2538  
FT 2539  
FT 2540  
FT 2541  
FT 2542  
FT 2543  
FT 2544  
FT 2545  
FT 2546  
FT 2547  
FT 2548  
FT 2549  
FT 2550  
FT 2551  
FT 2552  
FT 2553  
FT 2554  
FT 2555  
FT 2556  
FT 2557  
FT 2558  
FT 2559  
FT 2560  
FT 2561  
FT 2562  
FT 2563  
FT 2564  
FT 2565  
FT 2566  
FT 2567  
FT 2568  
FT 2569  
FT 2570  
FT 2571  
FT 2572  
FT 2573  
FT 2574  
FT 2575  
FT 2576  
FT 2577  
FT 2578  
FT 2579  
FT 2580  
FT 2581  
FT 2582  
FT 2583  
FT 2584  
FT 2585  
FT 2586  
FT 2587  
FT 2588  
FT 2589  
FT 2590  
FT 2591  
FT 2592  
FT 2593  
FT 2594  
FT 2595  
FT 2596  
FT 2597  
FT 2598  
FT 2599  
FT 2600  
FT 2601  
FT 2602  
FT 2603  
FT 2604  
FT 2605  
FT 2606  
FT 2607  
FT 2608  
FT 2609  
FT 2610  
FT 2611  
FT 2612  
FT 2613  
FT 2614  
FT 2615  
FT 2616  
FT 2617  
FT 2618  
FT 2619  
FT 2620  
FT 2621  
FT 2622  
FT 2623  
FT 2624  
FT 2625  
FT 2626  
FT 2627  
FT 2628  
FT 2629  
FT 2630  
FT 2631  
FT 2632  
FT 2633  
FT 2634  
FT 2635  
FT 2636  
FT 2637  
FT 2638  
FT 2639  
FT 2640  
FT 2641  
FT 2642  
FT 2643  
FT





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 8, 2004, 16:27:31 ; Search time 45 Seconds  
(without alignments)  
3828.288 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 2785  
Sequence: 1 MNQDKKNGAQAQSNPKSP.....APSTASGGTGPQBPFTSARA 546

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_25:\*  
1: sp\_archea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2785	100.0	546	Q86YW3	Q86YW3 homo sapien
2	2780	99.8	546	Q86T54	Q86T54 homo sapien
3	2777	99.7	546	Q86T86	Q86T86 homo sapien
4	2769	99.4	546	Q86T85	Q86T85 homo sapien
5	1294	46.5	524	Q8BHN1	Q8BHN1 mus musculu
6	1284.5	46.1	528	Q9NUQ3	Q9NUQ3 homo sapien
7	1276	45.8	718	Q86T52	Q86T52 homo sapien
8	1275	45.8	715	Q8N313	Q8N313 homo sapien
9	1244	44.7	676	Q91969	Q91969 gallus gall
10	1229.5	44.1	684	Q8VBT1	Q8VBT1 mus musculu
11	1192	42.8	463	Q8BUK2	Q8BUK2 mus musculu
12	1008	36.2	505	Q8N3S2	Q8N3S2 homo sapien
13	646	23.2	241	Q8BFI1	Q8BFI1 mus musculu
14	598	21.5	116	Q86V86	Q86V86 homo sapien
15	597.5	21.5	515	Q9VBL3	Q9VBL3 drosophila
16	503	18.1	264	Q86LQ1	Q86LQ1 branchiosto

17	462.5	16.6	436	10	Q84VE3	Q84VE3 oryza sativ
18	457	16.4	241	11	Q8BWJ7	Q8BWJ7 mus musculu
19	451.5	16.2	404	10	Q8RXD7	Q8RXD7 arabidopsis
20	442.5	15.9	186	4	Q9P0X1	Q9P0X1 homo sapien
21	400.5	14.4	150	11	Q8CSX1	Q8CSX1 mus musculu
22	372.5	13.4	335	5	Q22666	Q22666 caenorhabdi
23	369.5	13.3	181	4	Q8BZA4	Q8BZA4 homo sapien
24	336	12.1	372	10	Q9FGD8	Q9FGD8 arabidopsis
25	291	10.4	1168	5	Q9VYU0	Q9VYU0 drosophila
26	281	10.1	900	3	Q9P3P5	Q9P3P5 neurospora
27	281	10.1	992	4	Q9NTH6	Q9NTH6 homo sapien
28	278.5	10.0	1455	4	Q9UPV0	Q9UPV0 homo sapien
29	261	9.4	2033	10	Q7XEH4	Q7XEH4 oryza sativ
30	260	9.3	1937	6	Q9TV62	Q9TV62 sus scrofa
31	258	9.3	1206	11	Q08815	Q08815 rattus norv
32	258	9.3	1941	13	Q9DGM4	Q9DGM4 gallus gall
33	257.5	9.2	610	11	Q9CS72	Q9CS72 mus musculu
34	257	9.2	1003	12	Q91LX9	Q91LX9 kaposi's sa
35	257	9.2	1944	13	Q9DGM5	Q9DGM5 gallus gall
36	255	9.2	1943	13	Q8JG72	Q8JG72 gallus gall
37	254.5	9.1	609	17	Q8TXA4	Q8TXA4 methanopyru
38	254.5	9.1	1089	12	Q40947	Q40947 kaposi's sa
39	254	9.1	1939	13	Q9PTY2	Q9PTY2 gallus gall
40	251.5	9.0	728	4	Q60561	Q60561 homo sapien
41	251.5	9.0	1005	4	Q9P216	Q9P216 homo sapien
42	251	9.0	1940	13	Q8AY28	Q8AY28 gallus gall
43	250.5	9.0	1233	11	Q54988	Q54988 mus musculu
44	250.5	9.0	2139	5	Q07569	Q07569 entamoeba h
45	250	9.0	1001	11	Q88664	Q88664 rattus norv

#### ALIGNMENTS

#### RESULT 1

Q86YW3	ID	Q86YW3	PRELIMINARY;	PRT;	546 AA.
AC	Q86YW3;				
DT	01-JUN-2003	(TREMELrel. 24, Created)			
DT	01-JUN-2003	(TREMELrel. 24, Last sequence update)			
DT	01-JUN-2003	(TREMELrel. 24, Last annotation update)			
DE	Taxilin.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RX	MEDLINE=22448667; PubMed=12558796;				
RA	Nogami S., Satoh S., Nakano M., Shimizu H., Fukushima H., Maruyama A.,				
RA	Terano A., Shirataki H.;				
RT	"Taxilin; a novel syntaxin-binding protein that is involved in Ca2+-				
RT	dependent exocytosis in neuroendocrine cells.";				
RL	Genes Cells 8:17-28(2003).				
DR	EMBL; AF516206; AAC42465.1; ..				
SQ	SEQUENCE 546 AA; 61891 MW; 698CD74F78897DF6 CRC64;				

Query Match 100.0%; Score 2785; DB 4; Length 546;  
Best Local Similarity 100.0%; Pred. No. 1.1e-124;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKNQDKKNGAQAQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSGSQAPKPKPGAQA	60
Db	1	MKNQDKKNGAQAQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSGSQAPKPKPGAQA	60
Qy	61	RTAQSGALRDVSEELSRQLEDILSTVCVNNQGGCGEDCAQGEPAEPEDAEKSRITYARN	120
Db	61	RTAQSGALRDVSEELSRQLEDILSTVCVNNQGGCGEDCAQGEPAEPEDAEKSRITYARN	120
Qy	121	GSPEPTPVVNGKEPKSGKDPNTEINQSGVGDGRHRRPOEKKAKGLGKEITLLMOTLN	180
Db	121	GSPEPTPVVNGKEPKSGKDPNTEINQSGVGDGRHRRPOEKKAKGLGKEITLLMOTLN	180

```

181 TLSTPEEKLAALCKKYAELLSEHRNSQKMLLOKQSQLVQVQVTLNDIQLQMEQHNRNSKLRQ 240
181 TLSTPEEKLAALCKKYAELLSEHRNSQKMLLOKQSQLVQVQVTLNDIQLQMEQHNRNSKLRQ 240
241 LESLCRELQRHNSLRKEEGVQARAREEKKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300
241 LESLCRELQRHNSLRKEEGVQARAREEKKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300
301 ENNELAERLKKLISQYELREBEHIDKVFHKLQOQQLVDKLOQAQEMLKAEERHOREKD 360
301 ENNELAERLKKLISQYELREBEHIDKVFHKLQOQQLVDKLOQAQEMLKAEERHOREKD 360
361 FLKKEAVESQRMCELMKQOETHLKKQALYATEKFEFQNTLSKSEVFTTFKQEMEKMTK 420
361 FLKKEAVESQRMCELMKQOETHLKKQALYATEKFEFQNTLSKSEVFTTFKQEMEKMTK 420
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVRDKEGLQVKIORLEKLCRALQTERND 480
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVRDKEGLQVKIORLEKLCRALQTERND 480
481 LNKRVQDLSAGQSSLTDSGPERPEGPGQAQAPSPRVTEAPCPYGPAPSTASQGTGPQE 540
481 LNKRVQDLSAGQSSLTDSGPERPEGPGQAQAPSPRVTEAPCPYGPAPSTASQGTGPQE 540
541 PTSARA 546
541 PTSARA 546

RESULT 2
Q86T54 PRELIMINARY; PRT; 546 AA.
ID Q86T54
AC Q86T54;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein DKEZP451K215.
GN DKEZP451K215.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Koerner K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RA Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832338; CAD91138.1; -
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61861 MW; 728D674F78897DF6 CRC64;

Query Match 99.8%; Score 2780; DB 4; Length 546;
Best Local Similarity 99.8%; Pred. No. 1.8e-124;
Matches 545; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKEGAQA 60
DB 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKEGAQA 60
QY 61 RTAQSGALRDVSELSRQLEDILSTYCVDDNQGPGEDGAQGEPAEPDAEKSTTYARN 120
DB 61 RTAQSGALRDVSELSRQLEDILSTYCVDDNQGPGEDGAQGEPAEPDAEKSTTYARN 120
QY 121 GEPTPTPVNGEKPSPKDPNTEIRQSDVGDGRDRRPPQKKAKGLGKEITLLMQTLN 180
DB 121 GEPTPTPVNGEKPSPKDPNTEIRQSDVGDGRDRRPPQKKAKGLGKEITLLMQTLN 180
QY 181 TLSTPEEKLAALCKKYAELLSEHRNSQKMLLOKQSQLVQVQVTLNDIQLQMEQHNRNSKLRQ 240
DB 181 TLSTPEEKLAALCKKYAELLSEHRNSQKMLLOKQSQLVQVQVTLNDIQLQMEQHNRNSKLRQ 240
181 TLSTPEEKLAALCKKYAELLSEHRNSQKMLLOKQSQLVQVQVTLNDIQLQMEQHNRNSKLRQ 240
181 TLSTPEEKLAALCKKYAELLSEHRNSQKMLLOKQSQLVQVQVTLNDIQLQMEQHNRNSKLRQ 240

```

```

241 LESLCRELQRHNSLRKEEGVQARAREEKKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300
241 LESLCRELQRHNSLRKEEGVQARAREEKKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300
301 ENNELAERLKKLISQYELREBEHIDKVFHKLQOQQLVDKLOQAQEMLKAEERHOREKD 360
301 ENNELAERLKKLISQYELREBEHIDKVFHKLQOQQLVDKLOQAQEMLKAEERHOREKD 360
361 FLKKEAVESQRMCELMKQOETHLKKQALYATEKFEFQNTLSKSEVFTTFKQEMEKMTK 420
361 FLKKEAVESQRMCELMKQOETHLKKQALYATEKFEFQNTLSKSEVFTTFKQEMEKMTK 420
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVRDKEGLQVKIORLEKLCRALQTERND 480
421 KIKKLEKETTMYRSRWESSNKALLEMAEKTVRDKEGLQVKIORLEKLCRALQTERND 480
481 LNKRVQDLSAGQSSLTDSGPERPEGPGQAQAPSPRVTEAPCPYGPAPSTASQGTGPQE 540
481 LNKRVQDLSAGQSSLTDSGPERPEGPGQAQAPSPRVTEAPCPYGPAPSTASQGTGPQE 540
541 PTSARA 546
541 PTSARA 546

RESULT 3
Q86T86 PRELIMINARY; PRT; 546 AA.
ID Q86T86
AC Q86T86;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Hypothetical protein DKEZP451I0918.
GN DKEZP451I0918.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ausorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RA Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832636; CAD89951.1; -
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61892 MW; FEBCAD4F753F7FF1 CRC64;

Query Match 99.7%; Score 2777; DB 4; Length 546;
Best Local Similarity 99.6%; Pred. No. 2.5e-124;
Matches 544; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKEGAQA 60
DB 1 MKNQDKKNGAAKQSNPKSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKEGAQA 60
QY 61 RTAQSGALRDVSELSRQLEDILSTYCVDDNQGPGEDGAQGEPAEPDAEKSTTYARN 120
DB 61 RTAQSGALRDVSELSRQLEDILSTYCVDDNQGPGEDGAQGEPAEPDAEKSTTYARN 120
QY 121 GEPTPTPVNGEKPSPKDPNTEIRQSDVGDGRDRRPPQKKAKGLGKEITLLMQTLN 180
DB 121 GEPTPTPVNGEKPSPKDPNTEIRQSDVGDGRDRRPPQKKAKGLGKEITLLMQTLN 180
QY 181 TLSTPEEKLAALCKKYAELLSEHRNSQKMLLOKQSQLVQVQVTLNDIQLQMEQHNRNSKLRQ 240
DB 181 TLSTPEEKLAALCKKYAELLSEHRNSQKMLLOKQSQLVQVQVTLNDIQLQMEQHNRNSKLRQ 240
241 LESLCRELQRHNSLRKEEGVQARAREEKKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300
241 LESLCRELQRHNSLRKEEGVQARAREEKKKEVTSHFQVTLNDIQLQMEQHNRNSKLRQ 300
301 ENNELAERLKKLISQYELREBEHIDKVFHKLQOQQLVDKLOQAQEMLKAEERHOREKD 360

```

```

Db 301 ENWELAEERLKKLIEQYELREBEHIDKVFHKDLQOQLVDAKLOQAQEMLKEAERHQREKD 360
Qy 361 FLKKEAVESQRMCELMKQOETHLKQALALYTKFEFQNTLSKSSVFPTTFQOEMEKMXTK 420
Db 361 FLKKEAVESQRMCELMKQOETHLKQALALYTKFEFQNTLSKSSVFPTTFQOEMEKMXTK 420
Qy 421 KIKKLEKETTYMRSRWESSNKALLEMAEKTVDKXELGLEQLQVKIQLEKLCRALQTERND 480
Db 421 KIKKLEKETTYMRSRWESSNKALLEMAEKTVDKXELGLEQLQVKIQLEKLCRALQTERND 480
Qy 481 LNKRVODLSAGOGSLTDSGPERPFGCAAPSSPRVTEAPCYCAPSTASGQTGPQE 540
Db 481 LNKRVODLSAGOGSLTDSGPERPFGCAAPSSPRVTEAPCYCAPSTASGQTGPQE 540
Qy 541 PTSARA 546
Db 541 PTSARA 546

RESULT 4
Q86T85 PRELIMINARY; PRT; 546 AA.
AC Q86T85;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein DKFZP451J0118.
GN DKFZP451J0118.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ansorge W., Krieger S., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832637; CAD89952.1; -
KW Hypothetical protein.
SQ SEQUENCE 546 AA; 61846 MW; 36718BAE3A7B6C2 CRC64;

```

Query Match 99.4%; Score 2769; DB 4; Length 546;  
 Best Local Similarity 99.5%; Pred. No. 6.1e-124;  
 Matches 543; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 MNQDKNKGAQKSNPKSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQA 60
Db 1 MNQDKNKGAQKSNPKSSPGQPEAGPEGAQERPSQAAPAVEAGPGSSQAPRKPEGAQA 60
Qy 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNOGGPGEDGAGQGEPAEPEDAERSRTTVARN 120
Db 61 RTAQSGALRDVSEELSRQLEDILSTYCVDNNOGGPGEDGAGQGEPAEPEDAERSRTTVARN 120
Qy 121 GEPEPTPVNGKEPKSGDPNTEIRQSEVGDGRDHRPQKKGKAGLKGKITLLMQTLN 180
Db 121 GEPEPTPVNGKEPKSGDPNTEIRQSEVGDGRDHRPQKKGKAGLKGKITLLMQTLN 180
Qy 181 TLSTPEKLAALCKKYAEELREHNSOKMQLQKOSQVQKDHLEHSEKAVLARSK 240
Db 181 TLSTPEKLAALCKKYAEELREHNSOKMQLQKOSQVQKDHLEHSEKAVLARSK 240
Qy 241 LESLCELRQHRNSLKEEGVQARAREEKKRKYVTSHPQVTLNDIQLOMEQHNRNSKLQR 300
Db 241 LESLCELRQHRNSLKEEGVQARAREEKKRKYVTSHPQVTLNDIQLOMEQHNRNSKLQR 300
Qy 301 ENWELAEERLKKLIEQYELREBEHIDKVFHKDLQOQLVDAKLOQAQEMLKEAERHQREKD 360
Db 301 ENWELAEERLKKLIEQYELREBEHIDKVFHKDLQOQLVDAKLOQAQEMLKEAERHQREKD 360
Qy 361 FLKKEAVESQRMCELMKQOETHLKQALALYTKFEFQNTLSKSSVFPTTFQOEMEKMXTK 420
Db 361 FLKKEAVESQRMCELMKQOETHLKQALALYTKFEFQNTLSKSSVFPTTFQOEMEKMXTK 420

```

```

Qy 421 KIKKLEKETTYMRSRWESSNKALLEMAEKTVDKXELGLEQLQVKIQLEKLCRALQTERND 480
Db 421 KIKKLEKETTYMRSRWESSNKALLEMAEKTVDKXELGLEQLQVKIQLEKLCRALQTERND 480
Qy 481 LNKRVODLSAGOGSLTDSGPERPFGCAAPSSPRVTEAPCYCAPSTASGQTGPQE 540
Db 481 LNKRVODLSAGOGSLTDSGPERPFGCAAPSSPRVTEAPCYCAPSTASGQTGPQE 540
Qy 541 PTSARA 546
Db 541 PTSARA 546

RESULT 5
Q86HN1 PRELIMINARY; PRT; 524 AA.
AC Q86HN1;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RBBP7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Cortex, Head, and Testis;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK030100; BAC26785.1; -
DR EMBL; AK031783; BAC27547.1; -
DR EMBL; AK044130; BAC31791.1; -
DR MGD; MGI:1194910; Rbbp7.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0016564; P:transcriptional repressor activity; IDA.
DR GO; GO:0000122; P:negative regulation of transcription from P. .; IDA.
KW Hypothetical protein.
SQ SEQUENCE 524 AA; 60308 MW; 022877633B4ED7C CRC64;

```

Query Match 46.5%; Score 1294; DB 11; Length 524;  
 Best Local Similarity 53.9%; Pred. No. 5.2e-54;  
 Matches 265; Conservative 80; Mismatches 89; Indels 58; Gaps 5;

```

Qy 7 KNGAQSNNPKSSPGQPEAGPE---GAQERPSQAAPAVEAGPGSSQAP---RKPEGAQ 59
Db 12 RGGTTEASEGGRGRRRSPQKTEIGTWEEARICGLGVKADWVCNSQANDILQHDPSFC 71
Qy 60 ARTAQSGALR--DVSEELSRQLEDILSTYCV-DNNQGGPGEDGAGQGEPAEPEDAERSRTY 116
Db 72 GGTTKHSLGDEGSDFTTKRNLVSSVFTQEKREETPGREARTGPPDGGQDSECSR-- 129
Qy 117 VARNGEPEPTPVNGKEPKSGDPNTEIRQSEVGDGRDHRPQKKGKAGLKGKITLLM 176
Db 130 -----NKEKTLKEVLLLM 143
Qy 177 QTALNTLSTPEKLAALCKKYAEELREHNSOKMQLQKOSQVQKDHLEHSEKAVL 236
Db 144 QALNTLSTPEKLAALCKKYADLLEESRNVQKMKILQKQAVIKERHVLQSEHSEKAIL 203
Qy 237 ARSKLESICRLQHRNSLKEEGVQARAREEKKRKYVTSHPQVTLNDIQLOMEQHNRNS 296
Db 204 ARSKLESICRLQHRNKTKEENMQAREEERKEATAHFQITLNEIQALEQHDHNA 263
Qy 297 KLRQENMELAEERLKKLIEQYELREBEHIDKVFHKDLQOQLVDAKLOQAQEMLKEAERHQ 356
Db 264 KLRQENIELGSKLKKLIEQYALREBEHIDKVFHKELQOQLVDAKLOQTTLIKEADERHQ 323

```

```
QY 357 REKDFLLKEAVESQRMCELMKQOETHLKOOLALYTEKPEFQNTLSKSEVFTTFKQEME 416
DQ 324 REREFFLLKEATEBSHRKYEQOMQOEVQLKOOLSYMDKFEFQNTMAKSNELFPTTFQEME 383
QY 417 KMTKKIKKLEKETMTYRSRSESSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQOT 476
DQ 384 KMTKKIKKLEKETIIRWTKWNNKALLQWAEKTVDRKYEKAFQIKLELEKLCRALQOT 443
QY 477 ERNDLNKRVQDL 488
DQ 444 ERNELNEKVEVL 455

RESULT 6
Q9NUQ3 PRELIMINARY; PRT; 528 AA.
ID Q9NUQ3
AC Q9NUQ3
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE Hypothetical protein FLJ11209.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Watanabe S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Matsumoto J., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT *NEO human cDNA sequencing project.*;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK020711; BAA92068.1;
DR Genbank; HGNC:18578; CXorf15.
KW Hypothetical protein.
SQ SEQUENCE 528 AA; 60605 MW; CAE8D781DE06ACB3 CRC64;
```

```
Query Match 46.1%; Score 1284.5; DB 4; Length 528;
Best Local Similarity 53.4%; Pred. No. 1.5e-53;
Matches 285; Conservative 75; Mismatches 131; Indels 43; Gaps 10;

QY 28 EGAQRPQAPAEVAGPQSSQAPRK--PEGAQARTAGSGLRDVSELSRQLEDILT 85
DQ 7 EAARGGGGAEEATEAGRGRRSPQKPEIGTWERAGICGLGVKADMLCNSQSDILQH 66
QY 86 YCVDRNQGPGDGAQGEPAEPDAEKSRITYVARNGEPEPTPVNGEKEPSKGDPTNTEI 145
DQ 67 Q--GSNCGGTSNKHSLFEDSGSDFITENRNLS-----PAYCTQESRBEIPGG---BA 114
QY 146 RQSDVDRDRHRPQEKKAQKAGKKEITLIMQTLNLTSTPEEKLAALCKKYAEELLEHRN 205
DQ 115 RTDPPDQDSECNRNKEKT--LGKEVLLMQALNTLSTPEEKLAALCKKYADLLEERS 172
QY 206 SQQMKLQKQQLVQVQKHLEHSGSKAVIARSKLESLCRELQNRNRLKEGVQVARE 265
DQ 173 VQKQKILQKQQAQVKEKVKHLSHESKAILARSKLESLCRELQNRNRLKEGVQVARE 232
QY 266 EEBKREKVTSHFQVTLNDILOMEQHNRNSKLQENMELAEKILKILQVYELREEHDK 325
DQ 233 EEBRKEATNFIITLDEIQALQEQHDIENAKLRQENIELGEKILKILQVYALREEHDK 292
QY 326 VFQKHQLOOLVDAKLOQQAQEMKKEAEERHQRKDFLLKEAVESQRMCELMKQOETHLQ 385
DQ 293 VFQKHQLOOLVDAKLOQQAQEMKKEAEERHQRKDFLLKEAVESQRMCELMKQOETHLQ 352
QY 386 QLALYTEKPEFQNTLSKSEVFTTFKQEMKTKIKLEKETTYVRSRWSSNKALLE 445
DQ 353 QLSLYMDKPEEFQNTMAKSNELFPTTFQEMKTKIKLEKETTYVRSRWSSNKALLE 412
```

```
QY 446 MABEKTVDKLEGLQVKIQRLEKLCRALQOTERNDLNKRQV-----DL 488
DQ 413 MABEKTVDKLEGLQVKIQRLEKLCRALQOTERNELNEKVEVLKQVSIKAAKANDRL 472
QY 489 SAGGQGSUT--DSGPE-----RRPQGAQA--PSSPR--VTEAPCYPGAPSTEA 532
DQ 473 ATPVWQPCALDSHKELTSSKRGALGALEAPKQSRSAVQKPPSTGSAPIES 526

RESULT 7
Q86T52 PRELIMINARY; PRT; 718 AA.
ID Q86T52
AC Q86T52
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DE Hypothetical protein DKFZP451G083 (Fragment).
GN DKFZP451G083.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wisemann S.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL831988; CAD91140.1;
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 718 AA; 80290 MW; 16C3BF22C19559A4 CRC64;
```

```
Query Match 45.8%; Score 1276; DB 4; Length 718;
Best Local Similarity 45.9%; Pred. No. 5.1e-53;
Matches 277; Conservative 94; Mismatches 141; Indels 92; Gaps 10;

QY 3 NODKNGAAKQSNPKSPGQPEAGP--EGAQRPSQAAPAVEAEGPGSSQAPRKEGQA 60
DQ 38 NHSEQLSAEROSTP---PGDSSSLPSHNGLEKDGQ-----DSPTFVQPPER----- 81
QY 61 RPAQSGALRDVSELSRQLEDILTSTYCVDRNQGPGDGAQGEPAEPDAEKSRITYVARN 120
DQ 82 ---EASVHPDISELNRQLEDIINTY--GSAASTAGKGSARASEOPENASPDN--ED 133
QY 121 GEPEPTPVNGEKEPSKGDPTNTEIQRQSDVDRHRPQEKKAQKAGKIGKILTMQTLN 180
DQ 134 GDCEBTTBEAGREPVASGEPTVK---EPVSNKEQK--LEKCKILKGLGKEANLLMQNLN 187
QY 181 TLSTPEEKLAALCKKYAEELLEHRNSQKMKLQKQSQOLVQEKDHLRGEHSKAVLARSK 240
DQ 188 KLQTPPEEKDFLFRKYASLDEHRTQKGLKLLQKQVQIQEKDQLQGEHSRALLARSK 247
QY 241 LESLCRELQNRNRLKEGVQVAREEERKEKREKVTSHFQVTLNDILOMEQHNRNSKLQ 300
DQ 248 LESLCRELQNRNRLKEGVQVAREEERKEKREKVTSHFQVTLNDILOMEQHNRNSKLQ 307
QY 301 ENMELAEKILKILQVYELREEHDKVFKHQLQLOOLVDAKLOQQAQEMKKEAEERHQRK 360
DQ 308 ENTELAELKSIIDQVYELREEHDKVFKHQLQLOOLVDAKLOQQAQEMKKEAEERHQRK 367
QY 361 FLKKEAVESQRMCELMKQOETHLKOALYTEKPEFQNTLSKSEVFTTFKQEMKTKIK 420
DQ 368 YLLNQAAEWKILQAKMLKEQETVQLQALTYLSGRFEFQSTLTKSNEVATFKQEMDKTK 427
QY 421 KIKLEKETTYVRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQOTERN 480
DQ 428 NMKLEKETTYVRSRWSSNKALLEMAEKTVDKLEGLQVKIQRLEKLCRALQOTERN 487
QY 481 LNKRVQDL 488
DQ 488 LKKIRDAEISEKDDQSQHNSDRPEPESVNVQDEIDAEVNSVQTVAKNLATAFMIIEHP 547
```

```
QY 508 -----PGCAQAPSSP-----RVTEAPCYPCAPSTASGOTGPOE 540
Db 548 ESTPHQSKETQPEIGSSQESADAAALKEPEQPLIPSRDSESPPLPTPOAEAGGSDAEP 607
QY 541 PTSA 544
Db 608 PSKA 611

RESULT 8
Q8N3L3
ID Q8N3L3 PRELIMINARY; PRT; 715 AA.
AC Q8N3L3;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein DKF2p451f022 (fragment).
GN DKF2p451f022.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Fobo G., Han M.,
RA Osanger A., Wiemann S.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834248; CAD38924.2; -
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 715 AA; 79798 MW; 11C50191BCD26582 CRC64;

Query Match 45.8%; Score 1275; DB 4; Length 715;
Best Local Similarity 45.9%; Pred. No. 5.7e-53;
Matches 277; Conservative 94; Mismatches 141; Indels 92; Gaps 10;

QY 3 NODKNGAALQSNPKSPGQPEAGP--EQAQERPSQAAPVAEAGPGSSQAQPRKPEGAQA 60
Db 35 NHBSQLSAERQSTP---PGDSSSLPSHNGLEKEDGQ-----DSPTVPQPEK----- 78
QY 61 RTAQSGALDVSEELSRQLEDILSTVCVNNQGGPGEAGQEPAPPEAEPDAEKSRTTVARN 120
Db 79 ---EASVHPDISELNROLEDIINTY--GSAASTAGKEGSRASASEQENAESPDN---ED 130
QY 121 GEPEPTPVNGEKEPSKGDPTTEIRQSDVGDGRDHRPQEKKAQGLGKEITLLMOTLNL 180
Db 131 GDCEETTEAGRPVAGGEPTVK---EPVSNKEQK--LEKKILKGLGKEANLLMQLN 184
QY 181 TLSTPEKLAALCKKYAELLERHNSQKQMLQKQSQQLVQEKDHLRGEHSAVLARSK 240
Db 185 KLQTPREKFDLPFKYAEALDEHRTQKLLKLLQKQVQIQEKDQLQGEHSAVLARSK 244
QY 241 LESLCELOHNSLSKEEGVQARAEERKEKVTSHFOVTLNDIOLQMEQHNSRNSKLQ 300
Db 245 LESLCELOHNSLKEEALQAREERKEKVTSHFOVTLNDIOLQMEQHNSRNSKLQ 304
QY 301 ENNELAERLKLIEQVELREHIDKVPKHDQQLVDAKLOQAQEMLEKAEERHOREKD 360
Db 305 ENTELAEKLSIIOQVELREHLDKIFKREHQQKLVDAKLEQSQEMKAEERHOREKD 364
QY 361 FLKEAVESQRCMELKQOETHLQKQALYTEKPEEFQNTLSKSEVFTFKQEMEKWK 420
Db 365 YLLNQAEWLKQAVLKQETVLQAOQTLTYSRPFEPQSTLTNSNEVFATFKQEMDKTK 424
QY 421 KIKLKEKTTWRSRWSNKKALLEMAEKTVRDKEGLQVQIKRLEKLCRALQTERND 480
Db 425 KMKLEKDTATKARFENCNKALDMLBEKALRAKBYECFVKIGLENLCRALQERNE 484
QY 481 INKRVQLSAGGQGSITDSQERRPEG----- 507
Db 485 LHKKIRDAEISSEKDDQSQHNSDEEFSNVSVDQIDAEVNSVQTAVKNLATAFMIHHP 544
```

```
QY 508 -----PGCAQAPSSP-----RVTEAPCYPCAPSTASGOTGPOE 540
Db 545 ESTPHQSKETQPEIGSSQESADAAALKEPEQPLIPSRDSESPPLPTPOAEAGGSDAEP 604
QY 541 PTSA 544
Db 605 PSKA 608

RESULT 9
Q9I969
ID Q9I969 PRELIMINARY; PRT; 676 AA.
AC Q9I969;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Muscle derived protein.
GN MDP77.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Muscle;
RX MEDLINE=20175243; PubMed=10708594;
RA Uyeda A., Fukui I., Fujimori K., Kiyosue K., Nishimune H., Kasai M.,
RA Taguchi T.;
RT "MDP77: A novel neurite-outgrowth-promoting protein predominantly
RT expressed in chick muscles.";
RL Biochem. Biophys. Res. Commun. 269:564-569(2000).
DR EMBL; D89999; BAA94755.1; -
DR PIR; JC7222; JC7222.
SQ SEQUENCE 676 AA; 77020 MW; FCEA9E393250EE94 CRC64;

Query Match 44.7%; Score 1244; DB 13; Length 676;
Best Local Similarity 54.3%; Pred. No. 1.6e-51;
Matches 260; Conservative 81; Mismatches 104; Indels 34; Gaps 8;

QY 12 KQSNPKSPGQPEAGPQERPSQAAPVAEAGPGSSQAQPRKPEGAQAQTAQSGALRDV 71
Db 9 KQCVTTFTQDNQG-----QSKAEVFSQPLSP-TNQTSAQPEWATC-----DI 53
QY 72 SHELGRQLEDILSTVCVNNQGGPGEAGQEPAPPEAEPDAEKSRTTVARNGEPETP 127
Db 54 SHELNRQLEDIITKY-----GSAASLVEKEGTTATDTPKPEKEDVGSMEDEACED-- 102
QY 128 VNGEKEPSKGDPTTEIRQSDVGDGRDHRPQEKKAQGLGKEITLLMOTLNLSTPEE 187
Db 103 -VNESEKDKPAG-DASRAKPSKQKQ--LEKKILKGLGKEATLLMQSLNKLTPPE 158
QY 188 KLAALCKKYAELLERHNSQKQMLQKQSQQLVQEKDHLRGEHSAVLARSKLSLCE 247
Db 159 KLDDLFPKYAELLERHNSQKQMLQKQSQQLVQEKDHLRGEHSAVLARSKLSLCE 218
QY 248 IQHNSLSKEEGVQARAEERKEKVTSHFOVTLNDIOLQMEQHNSRNSKLQENNELAE 307
Db 219 LQHNKTLKBEITQAREBEKKEITNHFQSTLSEIQAIQEQSSRNKMLQOENTELAE 278
QY 308 RLKKLIEQVELREHIDKVPKHDQQLVDAKLOQAQEMLEKAEERHOREKDQFLKEAV 367
Db 279 KLSIIOQVELREHLDKIFKREHQQKLVDAKLEQSQEMKAEERHOREKKEEYLLNQAA 338
QY 368 ESQRCMELKQOETHLQKQALYTEKPEEFQNTLSKSEVFTFKQEMEKWKIKKLEK 427
Db 339 EWLQAKMLKEQSTVLQAOQTLTYSRPFEPQSTLTNSNEVFATFKQEMEKWKIKKLEK 398
QY 428 ETTWRSRWSNKKALLEMAEKTVRDKEGLQVQIKRLEKLCRALQTERNDLNRVQ 486
Db 399 DFTATWSRFPENCNRLDMLBEKALRAKBYECFVKIGLENLCRALQERNELYRKIK 457
```

	SEQUENCE FROM N.A.	
RP	STRAIN=Berkeley;	
RC	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,	
RA	George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutcliffe G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe W., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abriel J.F., Agayani A., An H.-J., Andrews-Flamknoch C., Baldwin D.,	
RA	Bailow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhargava D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,	
RA	de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Foster C., Gabrielsen A.C., Gary N.S., Gelbart W.M., Glasser K.,	
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,	
RA	Jaslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Milshina N.V., Mobarry C., Morris L.J., Moshrefi A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,	
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,	
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,	
RA	Spiet E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA	Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,	
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;	
RT	"The genome sequence of Drosophila melanogaster."	
RL	Science 287:2185-2195(2000).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Berkeley;	
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,	
RA	Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,	
RA	Gonzalez M., Guarini H., Li P., Liao G., Miranda A., Mungall C.J.,	
RA	Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,	
RA	Yu C., Lewis S.E., Rubin G.M., Celniker S.	
RL	Submitted (JUL-2001) to the ENBL/GenBank/DBJ databases.	
DR	EMBL; AE003753; AAF56538.1; --	
DR	EMBL; AF047502; AAK77234.1; --	
DR	FlyBase; FBgn0039379; CG5886.	
RSQ	SEQUENCE 515 AA; 59029 MW; 45B8F391D917094A CRC64;	
	Query Match 21.5%; Score 597.5; DB 5; Length 515;	
	Best Local Similarity 35.8%; Pred. No. 5.5e-21;	
	Matches 126; Conservative 98; Mismatches 105; Indels 23; Gaps 3	
OY	143 ZEIROSDVGDHRHRRPOEKKKAGLGEITLLMOTINTLTSTPEKLAALCKKYAELEE 202	
DB	2 EKLSKAKGVAREEVQRDKUEE-----LVMSLDECFSAEKVKU-----LQIR 45	
OY	203 HRNSOK-----QMVLKQKQSLOVQKDHLRGEISKAVIARSKLESLELRHORNSL 255	
DB	46 HVDSKNVSLRTAEILRVLQRWMSQQREKEQVORDLNKSVLMDKILQEVCREQRIKSV 105	
OY	256 KEEGVQVAREREERKGVTSHPVTINDIQLOMQEHNRNKLQENMELAEFLKLIEO 315	
DB	106 KNESLQIKVEERARKSQTFQPSLNDVORSLAKNNEINIKURDYNIETWKLIKLLAQ 165	
OY	316 YELREHHIDKVPKHUKLQQQLVDPAKLQAQAEMLKEAEERHQREKDFLKKEAVESORMCEL 375	

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: June 8, 2004, 16:37:57 ; Search time 50 Seconds  
(without alignments)  
3072.218 Million cell updates/sec

Title: US-10-023-529-44

Perfect score: 546

Sequence: 1 MNQDKKNGAAQSNPKSP.....APSTRASGQTGPQEPSTARA 546

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1155919 seqs, 281338677 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications\_AA.\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	546	100.0	546	9	US-09-976-740-44
2	546	100.0	546	12	US-10-023-529-44
3	546	100.0	546	13	US-10-023-529-44
4	546	100.0	546	13	US-10-023-529-44
5	546	100.0	546	15	US-10-023-529-44
6	476	87.2	530	12	US-10-276-774-2134
7	429	78.6	530	9	US-09-962-055-8
8	429	78.6	530	9	US-09-976-740-8
9	429	78.6	530	12	US-10-023-529-8
10	429	78.6	530	13	US-10-023-529-8
11	429	78.6	530	13	US-10-023-529-8
12	429	78.6	530	15	US-10-023-529-8
13	172	31.5	557	9	US-09-962-055-5
14	172	31.5	557	9	US-09-976-740-5
15	172	31.5	557	12	US-10-023-529-5

```
16 172 31.5 557 13 US-10-023-529-5 Sequence 5, Appli
17 172 31.5 557 13 US-10-023-529-5 Sequence 5, Appli
18 172 31.5 557 13 US-10-023-529-5 Sequence 5, Appli
19 17 3.1 85 15 US-10-023-529-5 Sequence 3413, Ap
20 11 2.0 436 12 US-10-023-529-5 Sequence 207434,
21 9 1.6 99 9 US-09-764-869-1082 Sequence 1082, Ap
22 9 1.6 99 14 US-10-091-504-1082 Sequence 1082, Ap
23 9 1.6 99 15 US-10-237-577-1082 Sequence 112, App
24 9 1.6 594 9 US-09-801-368-112 Sequence 1762, Ap
25 9 1.6 594 15 US-10-369-493-1762 Sequence 72115, A
26 9 1.6 637 12 US-10-282-122A-72115 Sequence 130, App
27 8 1.5 25 16 US-10-443-622-130 Sequence 48446, A
28 8 1.5 106 12 US-10-425-114-48446 Sequence 272074,
29 8 1.5 121 12 US-10-424-599-272074 Sequence 230732,
30 8 1.5 148 12 US-10-424-599-230732 Sequence 122, App
31 8 1.5 156 15 US-10-443-622-122 Sequence 316, App
32 8 1.5 229 14 US-10-080-170-316 Sequence 12433, A
33 8 1.5 413 14 US-10-156-761-12433 Sequence 21212, A
34 8 1.5 447 15 US-10-369-493-21212 Sequence 7865, Ap
35 8 1.5 450 12 US-10-335-977-7865 Sequence 7866, Ap
36 8 1.5 451 12 US-10-335-977-7866 Sequence 6430, Ap
37 8 1.5 484 9 US-09-738-626-6430 Sequence 23425, A
38 8 1.5 525 15 US-10-369-493-23425 Sequence 41, Appl
39 7 1.3 7 9 US-09-962-055-41 Sequence 41, Appl
40 7 1.3 7 9 US-09-976-740-41 Sequence 41, Appl
41 7 1.3 7 12 US-10-071-242-41 Sequence 41, Appl
42 7 1.3 7 13 US-10-023-529-41 Sequence 41, Appl
43 7 1.3 7 13 US-10-023-529-41 Sequence 41, Appl
44 7 1.3 7 15 US-10-616-187-41 Sequence 2071, Ap
45 7 1.3 18 14 US-10-225-567A-2071
```

## ALIGNMENTS

### RESULT 1

```
US-09-976-740-44
; Sequence 44, Application US/09976740
; Publication No. US20020194633A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/976,740
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-976-740-44
```

```
Query Match 100.0%; Score 546; DB 9; Length 546;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNQDKKNGAAQSNPKSPGPQEPGQAQRPQQAAPVAEAGPGSSQAPKPEGAQA 60
DB 1 MNQDKKNGAAQSNPKSPGPQEPGQAQRPQQAAPVAEAGPGSSQAPKPEGAQA 60
```